The Dissertation Committee for Gayatree Nandan Rao Certifies that this is the approved version of the following dissertation:

MEASURING PHONETIC CONVERGENCE: SEGMENTAL AND SUPRASEGMENTAL SPEECH ADAPTATIONS DURING NATIVE AND NON-NATIVE TALKER INTERACTIONS

Committee:

Randy L. Diehl, Co-Supervisor

Rajka Smiljanic, Co-Supervisor

Leslie B. Cohen

Lawrence K. Cormack

Zenzi M. Griffin

MEASURING PHONETIC CONVERGENCE: SEGMENTAL AND SUPRASEGMENTAL SPEECH ADAPTATIONS DURING NATIVE AND NON-NATIVE TALKER INTERACTIONS

by

Gayatree Nandan Rao, B.S.E; M.A.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

December, 2013

Dedication

To Naya, Riyaz and Ahilya.

Acknowledgements

I am grateful to Randy Diehl and Rajka Smiljanic for mentoring and guiding me through not just this dissertation but also a large part of my graduate career. Because of them, I have learned to be a curious and diligent researcher. Thanks also to the other committee members, Larry Cormack, Zenzi Griffin and Les Cohen for offering fresh perspectives on the dissertation. They were always available with willing ears and helpful feedback. And to Samantha Sebastian for her help with the pilot and initial planning of this dissertation. I owe a debt of gratitude to Bjorn Lindblom who showed me how fascinating and scientifically rigorous the study of phonetics could be. I would not be a phonetician had it not been for his teaching and influence.

I would like to express gratitude towards my friends in linguistics and psychology, specifically but certainly not limited to Britanny Hall-Clark, Anushka Pai, Cynthia Hansen, Kate Shaw Points, Jessica White Sustaita, Heeyoung Lyu, Taryne Hallet, Douglas Bigham, Dan Olson, Lori Czerwionka, Jordan Davison, Bharath Chandrasekaran, Sang-Hoon Park, Teresa York Morrison and Robin Fletcher. Without their love and support, this experience would have been meaningless. Special thanks to Kate, editor par excellence, who helped make this dissertation cleaner and clearer. I can't thank Kayla Cobb and Ashlee Willingham enough for watching my boy so I could work each day without having to worry about his well-being and safety.

I would also like to express my love and appreciation to my family by choice, Steven Hilderbrand, Kristen Harrison, Whitney Roberts, Chris Rauschuber, Sarah Rodriguez Pratt, David Pratt, Sirsha Chatterjee, Joey Martin and Amanda Field. They were always there with a supportive hug and a pint. And also to my family by blood and marriage, mama, papa, mom, dad (Reena Rao, Nandan Rao, Shaku Parikh, Ashok Parikh

respectively), Chi Sastry, Giri Sastry, Nina Thomas and Rome Thomas. Individually, you are all amazing and wonderful but together you are the best, most loving family anyone could ever ask for. Thank you for always believing in me. And Nina and Chitu, thank you for always keeping your phones nearby. Thanks to the littlest ones Naya, Ahilya and Riyaz for simply being perfect in every way. Riyaz, hugging you at the end of the day keeps me going. Between you and the dissertation, you are my favorite baby.

And finally, Sandeep Parikh, you are my partner and my rock. Your unwavering love and support make every experience and accomplishment worth so much more. Thank you for being on this journey with me. Let's go explore some more.

MEASURING PHONETIC CONVERGENCE: SEGMENTAL AND SUPRASEGMENTAL SPEECH ADAPTATIONS DURING NATIVE AND NON-NATIVE TALKER INTERACTIONS

Gayatree Nandan Rao, PhD

The University of Texas at Austin, 2013

Supervisors: Randy L. Diehl and Rajka Smiljanic

Phonetic convergence (PC) is speech specific accommodation characterized by an increase in similarity in a dyad's speech patterns due to an interaction. Previous research has demonstrated that PC occurs in dyads during various interactive tasks (*e.g.* map completion and picture matching) and in cross-linguistic conditions (*e.g.* dyads who speak the same or different native language) (Pardo, 2006; Kim et al., 2011). Studies suggest that speakers who are closer in linguistic distance (*i.e.* share the same native language) are more likely to converge than speakers who are far apart (*i.e.* speak different native languages) (Kim et al, 2011). However, Interdialectal conditions where speakers use different national dialects of the same language have been studied to a far lesser extent (Babel, 2010). Similarly, studies have examined both segmental and suprasegmental features that are susceptible to PC but rhythm has not been studied extensively (Krivokapic, 2013; Rao et al., 2011). Though initial studies postulated that PC is the result of either automatic or social processes, more current research suggests that a combination of both kinds of processes may be better able to account for PC (Goldinger, 1997; Shepard et al., 2001; Babel, 2009a).

The current dissertation uses novel measures such as Interlocutor Similarity and EMS + centroid to implicate global properties of vowels and rhythm respectively as acoustic correlates of PC. Moreover, it finds that speakers showed both convergence and divergence in vowels and rhythm as moderated by their language background. Close interactions between native speakers of American English (AE) resulted in convergence whereas interdialectal interactions (between AE and Indian English speakers) and mixed language interactions (between native and non-native speakers of AE who are native speakers of SP) resulted in both convergence and divergence. The results from this study may shed light on how speakers attenuate the highly variable nature of speech by adapting speech patterns to aid intelligibility and information sharing (Shepard et al., 2001) and that this attenuation is moderated by social demands such as identity and cultural distinctiveness.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AE: American English

AuE: Australian English

APS: Amplitude power spectrum

BE: British English

CI: Confidence Interval

EMS: Envelope modulation spectrum

F1: first formant

F2: second formant

IE: Indian English

IS: Interlocutor similarity

L1: A speaker's first or native language

L2: A speaker's second language

nPVI: normalized pairwise variability index

NZE: New Zealand English

PC: Phonetic Convergence

rPVI: raw pairwise variability index

SE: Standard Error

SP: Spanish

Spec-IS: specific IS

Sys-IS: systemic IS

Chapter 1: Introduction

Speech is an inherently collaborative and conscious act. Jakobson famously said, "We speak in order to be heard, in order to be understood" (Jakobson, Halle & Fant, 1952). Implicit in this statement is the presence of another person (or people) in the listener's role. In a dialogue, the ultimate purpose of speaking is to communicate effectively in order to achieve a common goal. This common goal could be a task that involves coordinating a hunting party, coordinating trips for groceries or simply trying to connect with another individual.

However, speech is not a simple act. One source of complexity, and potential difficulty for speech perception, lies in the inherent variability of the speech signal. Speech tokens vary within- and across- speakers due to differences in vocal tract size, speaking rate, socio-phonetic factors, such as dialect, and even biological factors such as vocal fold damage or having a cold. Phonetic convergence (PC) may be a result of the direct link between perception and production that aims to reduce this variability (Goldinger, 1998). As people interact, their behaviors such as body posture, word choice and speech become similar to each other. This increase in similarity takes place on both physiological and social levels. Such alignments are known by different terms such as accommodation, entrainment, and convergence. PC (also called phonetic imitation) is one specific component of this general accommodation that involves increasing similarity in the speech patterns of interlocutors. It has been hypothesized that PC may aid speech intelligibility through the reduction of variability (Pickering & Garrod, 2004). Furthermore, speech entrainment at various linguistic levels may aid communication and increase the chances of task success (Nenkova, Gravano & Hirshberg, 2008). PC is different from other goal-oriented speaking style adaptations, such as infant- and

computer-directed speech, clear speech and Lombard speech. The latter involve speakers adjusting their output to meet the perceptual demands of their target audience or the communicative situation by slowing speech rate or increasing perceptual distance between phonetic units presumably leading to improved speech recognition (Junqua 1993; Kuhl et al. 1997; Skowronski & Harris 2006). Unlike these modifications, PC is the result of listener-oriented adjustments to interlocutors' speech patterns.

Although the underlying mechanisms and the communicative and long-term consequences of these adaptations are not well understood, research suggests that PC can result either from automatic (Goldinger, 1998; Pickering & Garrod, 2004) or social (Shepard, Giles & Le Poire, 2001) processes. Convergence appears to be the expected default that can occur either due to automatic alignment in speech or due to social reasons such as the desire to present oneself favorably. However, people's behavior can also become more disparate during an interaction and social factors are necessary to motivate this divergence (Bourhis & Giles, 1977; Babel 2009a; Babel 2010). Convergence may be the result of durable changes to mental representations and can be generalized to novel tokens, suggesting that learning plays a role in PC (Delvaux & Soquet, 2007; Nielsen, 2008). PC has been demonstrated within speakers of varying linguistic backgrounds using tasks such as reproducing map routes, picture matching, speech shadowing and passive listening (Pardo, 2006; Krivokapic, 2013, Kim, Horton & Bradlow, 2011; Delvaux & Soquet, 2007).

Kim et al. (2011) define interlocutor linguistic distance based on a dyad's native language and dialect. Speakers who share the same native language (L1) and dialect are the closest in distance whereas those speakers that do not share the same L1 are considered the farthest. Speakers who share an L1 but do not share the same dialect (*e.g.* southern American English vs. northern American English) fall somewhere in the middle.

They found that dyads that are close in terms of linguistic distance are more likely to converge than dyads that are far apart. Kim et al.'s (2011) study did not deal with dyads that share an L1 but speak differing national varieties of it. However, research by Krivokapic (2013) and Babel (2010) has shown that speakers may be susceptible to convergence if they speak differing national dialects of a language. With the exception of Krivokapic (2013) and Rao, Smiljanic and Diehl, (2011c), PC in rhythm has been largely unexplored. Interactions between native and non-native speakers have also not been studied extensively (Kim et al., 2011; Lewandowski, 2012). These studies are discussed in greater detail in Chapter 2.

Keeping these gaps in mind, the main focus of investigation in this dissertation is the role of linguistic distance in a dyad's ability to converge or diverge or more generally, to adapt. Using terminology pertaining to the interlocutor language distance, the specific questions explored in the current study are:

1. Do speakers within a dyad adapt if they are close in linguistic distance (*i.e.* they share the same language and national dialect)?
2. Do speakers within a dyad adapt if they are further apart in linguistic distance (*i.e.* they speak different national dialects of the same language)?
3. Do speakers within a dyad adapt if they are linguistically far apart (*i.e.* one is a native speaker a target language and one is not)?

Using American English (AE) as the target language, the current study examines interactions in three language conditions: between L1 speakers of AE (linguistic distance: near), L1 speakers of AE and Indian English (IE) (linguistic distance: intermediate), and L1 speaker of AE and L2 speakers of AE who are L1 speakers of Spanish (linguistic distance: far). Kim et al., (2011) use intermediate to refer to dyads that share the same L1 but speak differing regional dialects. In order to keep the intermediate group separate

from speakers who speak differing national dialects of an L1, the second group in the current study was called the 'interdialectal' group. These three language conditions will be referred to as the native language, mixed dialect and mixed language groups respectively. These languages were selected for two reasons. Firstly, they share four of the six vowels that are examined in this study. AE, IE and Spanish (SP) vowel inventories overlap in /i, u, o, e/. Secondly, based on phonological descriptions, AE and SP are considered prototypically stress-timed and syllable-timed respectively, while IE is considered mixed. Even if phonological rhythm descriptions are debated, the rhythmic differences in these types of languages are readily perceptible (Ramus, Dupoux & Mehler, 2003; Ramus & Mehler, 1999) and provide a starting point in exploring rhythm convergence.

In the current study, phonetic convergence patterns are investigated at two levels: segmental (vowel space) and suprasegmental (rhythm). Vowels are examined in this study to serve as a 'control' of sorts. Since vowel convergence has been extensively explored, speakers adapting in the case of vowels but not in the case of rhythm would suggest that suprasegmental convergence is not as important to speakers as segmental convergence is. Alternatively, no vowel adaptations would signal an issue with the methodology itself. Furthermore, this study attempts to provide novel measures of vowels that track pairwise changes that take place to a dyad's vowels on a vowel specific and vowel systemic basis. Even though convergence involves alignment in segmental and suprasegmental properties of speech, few studies have explored convergence in suprasegmental properties of speech, especially rhythm (Krivokapic, 2013; Pardo, Jay & Krauss, 2010, Rao et al., 2011c; Lewandowski, 2012). This becomes even more crucial given Goldinger's (1998) proposal that temporal and melodic properties of speech may be particularly susceptible to convergence (p. 259). With that in mind, this dissertation

provides novel insights into PC through detailed and systematic examination of how global temporal speech properties, as measured by a spectral measure of rhythm, are affected by convergence. This research aims to demonstrate that:

1. Speakers accommodate to the suprasegmental (rhythm) and segmental (vowel) patterns of their partners during a spontaneous conversation task such as a map task.
2. Linguistic background moderates convergence in both vowels and rhythm.

In addition to examining PC for previously understudied linguistic backgrounds and suprasegmental adaptations, this dissertation is innovative in its use of novel methods to detect convergence. Vowels are examined using the traditional midpoint formants measurement as well as another approach called vowel internal spectral change (VISC). While VISC has been used extensively to study vowels, it has not been used to examine convergence. Both Peterson and Barney (1952) and Hillenbrand (2011) point out that single time point measurements are not sufficient to represent vowels accurately. VISC incorporates dynamic information and presents an alternative to steady state measurements of vowel convergence. Another novel measure of vowel convergence, called the interlocutor similarity (IS), is also proposed. This measure has two versions, specific IS and systemic IS. Systemic IS (sys-IS) is a measure of vowel convergence between a speaker's vowels and his or her partner's vowels considered as a set whereas specific IS (spec-IS) is a measure of the same dyad's vowel convergence considered on a per vowel basis. Both these measures combine a dyad's formant information into a straightforward similarity measure that can be tracked over time. Sys-IS tracks this similarity change on a systemic global basis whereas spec-IS does so on a more granular (per-vowel) basis.

Rhythm is also measured using a relatively novel spectral approach called envelope modulation spectrum (EMS). Unlike, Krivokapic (2013) who used a traditional duration measure to examine PC, the current study uses a spectral measure. Spectral measures circumvent certain pitfalls of traditional approaches by not making presuppositions based on phonological descriptions, including dysfluencies and filled pauses; additionally, these measures are not time-consuming to implement (Liss, LeGendre & Lotto 2010). Liss et al. (2010) stress that the key difference between EMS and duration metrics is that the former may be able to detect rhythmic information that goes beyond linguistic structure (*e.g.* rhythmic variability due to speech disorders such as dysarthria or individual speaker rhythm), making this measure ideal to track global rhythm changes in a dyad's speech pattern dependent on each speaker's idiosyncratic rhythm.

The findings from the current study show that speakers belonging to each language condition exhibit varying degrees of vowel-specific and vowel-general adaptations based on social and linguistic motivations. Specifically, vowel adaptations were noted across language conditions with marked vowels being more likely to diverge. Vowels were considered marked or salient if they were part of one speaker's repertoire but not the other. Rhythm adaptations were also noted in the groups that shared L1 (specifically, English) but not in the one that contained dyads of L1 and L2 speakers of English. This can be best explained via the rhythm continuum of stress-timed and syllable-timed languages; speakers who shared rhythmic properties (close and interdialectal groups) showed PC whereas those speakers that had linguistic rhythms lying at the ends of the continuum did not show any adaptations. The current study reveals vowel and rhythm convergence and divergence occurring simultaneously within the same set of dyads, supporting Babel's claim that both social and automatic theories

must be involved in convergence. Adaptation processes such as these are important for understanding speaker-listener interactions, plasticity in segmental and suprasegmental representations as well as factors responsible for language sound change and other aspects of sociolinguistic phonetic convergence such as social group and identity construction.

It is important to note that, dependent on the field of research, the terms convergence and divergence are often used to refer to processes that can vary vastly. For example, within sociolinguistics, Labov (2002) explains that convergence is expected in the languages of communities that are in communicative contact and divergence is expected in the languages of communities that are separated (*e.g.* geographically). Thus, in the context of language change and contact, convergence and divergence are used to refer to long-term linguistic changes that result in changes to the sound or syntactic structure or even social status of a language and can occur as a result of factors such as dialect contact or social factors (Labov, 1982; Labov, 2002).

The current study uses PC (or simply convergence or adaptation) to refer to spontaneous phonetic imitation: short-term linguistic changes occurring to an individual's speech during verbal interactions between interlocutors. For the purposes of this study, convergence will refer to short-term phonetic and phonological changes that occur to interlocutors' speech during verbal interaction. Since convergence and divergence are also used to denote language change on a long-term basis, such changes will be referred to explicitly as long-term convergence and divergence. It should be noted that the terms 'long-term' and 'short-term' refer to the time it takes for these changes to occur and do not predict how long these changes last.

Chapter 2 lays out the background literature for PC, rhythm and vowels. Chapter 3 elaborates on the methodology used in the current study and the rationale behind the

analyses employed in the study. This is also where the novel metrics, systemic IS, specific IS and EMS + centroid are described in detail. Chapters 4, 5 and 6 deal with the individual experiments of native language, mixed dialect and mixed language pairings respectively. In order to examine the controlled tokens separately, these chapters only examine the pre- and post-task tokens for vowels and rhythm. Chapter 7 examines the changes that took place to the speaker's rhythm during the map interaction itself. Chapter 8 evaluates the role of accent imitation on PC. Chapter 9 concludes with a general discussion of the findings from all experiments. Definitions of phonetic concepts that are relevant to the dissertation are provided in Appendix A.

Chapter 2: Background

This chapter provides the theoretical and empirical foundation for the current dissertation. It focuses on the background literature concerning PC, vowels and rhythm, in that order. The section on PC focuses on the historical background and the theories that have been used to explain the phenomenon. It also includes the methods and acoustic properties that have been used to examine PC. The section on rhythm also outlines the history of linguistic rhythm research including the traditional duration measures that have been used in rhythm classification. There is a discussion of the issues surrounding this research and descriptions of more recent spectral measures of rhythm. The penultimate section in the chapter has a brief description of vowels and the measures that have been used to quantify and classify them. This chapter concludes with a summary and evaluation of the most relevant points for the current study regarding the role of linguistic background on PC specifically with a focus on vowels and rhythm.

2.1 PHONETIC CONVERGENCE (PC)

2.1.1 Origin and history

Phonetic convergence is defined as an increase in a dyad's segmental and suprasegmental similarity (Pardo, 2006). Based on Bloomfield's (1935) principle of density (a speaker adapts his or her speech to that of an interlocutor) and Labov's (1966) study on long-term language change, Giles (1973) proposed a model for short-term convergence or PC. His theory of accent mobility states that all speakers have 'an accent repertoire' with two levels: primary and secondary. The primary level is a continuum of an accent where the standard pronunciation and regional dialect of a language lie on

opposite ends. A speaker can alter his or her speech such that his or her accent lies along this continuum. It is possible for a speaker to have more than one continuum (*e.g.* multilinguals). The secondary level comprises accents that a speaker can mimic but does not regularly use. This level is used if a speaker relocates to an area where the accent is foreign enough to be excluded from the primary level. It is possible for transference to occur so that accent information from the secondary level is assimilated into the primary level. The accent mobility model allows for accent change to result in convergence or divergence (both long-term and short-term). However, convergence can only occur when dissimilarities are reduced for social approval, whereas divergence occurs as a result of the speaker's need to dissociate from his or her partner. Thus, according to Giles (1973), both phonetic convergence and divergence are a result of social factors.

Currently, two main accounts of accommodation in speech have been proposed in the field of social psychology: automatic (Goldinger, 1997; Pickering & Garrod, 2004) and social (Shepard, Giles & Le Poire, 2001). An automatic cognitive process may occur with easy, highly familiar or learned tasks, takes little or no attention and can take place in parallel with other processes (Schneider & Shiffrin, 1977). This suggests that both the automatic and social theories of accommodation may be automatic from a cognitive perspective. Both approaches require little attention and take place in parallel with other tasks such as picture matching or completing map tasks. Automatic and social theories of accommodation are described in detail below.

2.1.2 Automatic theories of accommodation

Automatic theories of accommodation suggest that speech production is affected by perceptual details. The exemplar theory (Goldinger, 1998) proposed that stored

episodic memory details are the basis of accommodation by showing that that speakers tended to imitate words and non-words that they were asked to shadow (repeat after listening to a target word). In order to determine the effect of word frequency on imitation, shadowing reaction times (RT) were also tracked. The target tokens were submitted to the AXB perceptual judgments. In an AXB task, a separate set of participants listen to three tokens (A, B and X) and judge whether X sounds more like token A or B (also called an ABX task). The results demonstrated convergence in perception. Furthermore, this convergence was mediated by lexical factors. Low frequency words were imitated to a greater extent than high frequency words. He theorized that stronger episodic representations of higher frequency words yielded faster responses in shadowing RTs. However, because a more generic profile is created as a lexical representation for high frequency words, the imitative response is less or weaker than it would be for a low frequency word. Low frequency words are influenced by more specific and fewer traces in memory or weaker episodic representations, yielding stronger imitations. Although exemplar theories tend to be associated with lexical items, this study prompted Goldinger to suggest that perceptual details (phonemic properties or features such as vowel quality) may be stored along with traces of entire words and also be subject to analogous imitation patterns. Goldinger (1998) suggested that according to episodic theories, idiosyncratic aspects of speech such as voice details (which would be treated as noise and normalized during perception) are stored as perceptual information to be utilized in later perception.

Pickering & Garrod's (2004) interactive alignment model of dialogue processing is another instance of an automatic theory of accommodation. According to this model, interlocutors develop alignment at every level (*e.g.* syntactic or lexical) of linguistic representation during interactions. These levels include situational model, syntactic,

semantic, lexical, phonological and phonetic representations. For example alignment at the semantic level would involve using the same interpretation of a word once it has been introduced with that specific interpretation. This process is automatic and occurs due to priming. Moreover, there is feedback across all levels of linguistic representation leading to enhancement of alignment at one level from various levels. Thus, this theory assumes a direct and coupled relationship between production and perception of language. Within this framework, PC is alignment specifically at the phonetic level. The authors posit that alignment is necessary between interlocutors. Divergence only takes place in monologues. Misalignments should they occur, are repaired using common ground or joint knowledge.

2.1.3 Social theories of accommodation

Proponents of social theories of accommodation, on the other hand, state that interlocutors are capable of strategically manipulating speech to navigate social distance (Giles, 1973; Shepard, Giles & Le Poire, 2001). The most prominent social theory is called Communication Accommodation Theory (CAT), which describes language behavior within interpersonal and intergroup interactions. It states that accommodation is used to mark status within a group or set oneself in alignment or apart from another individual or group. There are four approximation strategies outlined in CAT: convergence, divergence, maintenance and complementary. In this view, convergence is used to align oneself with an interacting partner in accent, dialect, idiom or code-switching strategies. Divergence is used to distinguish oneself in the same aspects, often due to disagreement with the interlocutor. The authors suggest that convergence may be used to gain approval or to make the interaction as smooth as possible. Accommodation

can be unimodal or multimodal taking place across auditory, visual and spatial modalities. It can be symmetrical or asymmetrical, where one speaker tends to accommodate more than the other (Pardo, 2006). The last two strategies involve maintenance of and setup of a complementary accent (where distinctive features in one's voice or accent are emphasized). These strategies fall outside the scope of this paper and will not be described.

While the automatic account of speech accommodation suggests that speech production is affected by 'episodic aspects of lexical representation' (Goldinger, 1998) or priming (Pickering & Garrod's, 2004), CAT claims that speakers are able to manipulate convergence to mark one's own in-group and out-group status. Although automatic and social theories are often presented in opposition to each other, often factors from both kinds of theory are necessary to explain reasons for convergence. Speakers may negotiate automatically induced adaptations to create social group delineations between themselves and an interlocutor (Babel, 2009a, 2010; Giles, 1997). Social factors are explored further in Section 2.1.5.

Research within phonetics has focused on acoustic components of speech that are subject to convergence. A number of experimental setups such as shadowing, ambient listening paradigms and dyadic interactions have been used to examine PC (Goldinger, 1997; Goldinger, 1998; Pardo, 2006; Kim et al., 2011; Delvaux & Soquet, 2007). Some considerations in inducing and analyzing PC are described below in Sections 2.1.4.

2.1.4 Tasks that induce PC

PC has been demonstrated in a variety of tasks that can track speech adaptations either in one speaker or across a dyad. Studies have used word shadowing to elicit

convergence in subjects (Goldinger 1997, 1998; Babel 2009a). In these experiments, speakers were asked to repeat target tokens produced by a model speaker heard over earphones. Delvaux and Soquet (2007) used a similar paradigm to show convergence of speakers' speech to ambient speech (pre-recorded sentences) played over headphones or a speaker in the absence of repetition. Convergence has also been demonstrated in more interactive tasks such as map and diapix (Pardo, 2006; Kim, Horton, & Bradlow, 2011). In a map task, two speakers are given similar maps and paired as instruction provider (or simply provider) and instruction receiver (or receiver). The provider's map contains minor differences from the receiver's map and a route drawn around landmark items. The provider is tasked with giving verbal instructions that allow the receiver to reproduce the route drawn on the provider's map. For a picture matching or diapix task, each speaker pair is provided with pictures of everyday scenes that vary slightly (*e.g.* a day at the beach). Speakers are tasked with discussing their picture and noting similarities and dissimilarities with their speaking partner. Shadowing tasks allow for a more controlled set of data to be obtained because one speaker repeats a predetermined set of words or sentences. Interactive tasks, on the other hand, allow for a more conversational setting that obtains more free-form recordings. Detection of PC was achieved either using perception-based AXB tasks (Pardo, 2006; Kim et al., 2011) or production-based acoustical analyses (Delvaux and Soquet, 2007).

Initial studies used AXB listening tasks to show that listeners were sensitive to the changes in the speaker's pronunciation patterns, *i.e.*, that speakers' pronunciations were becoming more similar to a model speaker's or interaction partner's productions (Goldinger, 1997; Goldinger, 1998; Pardo, 2006; Kim et al., 2011). While these studies were useful in demonstrating convergence, they did not identify specific phonetic factors of speech that are subject to convergence and that give rise to the perceptual impressions

of speech pattern similarities. Subsequent research has utilized acoustic analyses to examine various segmental and suprasegmental features as possible targets of convergence such as VOT (voice onset time, difference between the release of the oral closure and the start of the vocal fold vibration) (Nielsen, 2008; Sancier & Fowler, 1997), Mel-frequency cepstral coefficients (MFCC, power cepstrum that approximates the human auditory system response) of vowels and consonants (Delvaux & Soquet, 2007), articulation rates (Pardo et al., 2010) and stress (Chiosáin, 2007).

2.1.4.1 Segmental correlates of PC

Most of the research conducted on phonetic convergence has focused on segmental features. Delvaux and Soquet (2007) showed in two separate studies that both vowels and consonants undergo convergence. Both of these studies examined female speakers of the Mons and Liège dialects of Belgian French when exposed to ambient recordings of the dialect that they did not speak. Participants read aloud sentences before and after hearing model speakers from the dialect other than their own speak a set of sentences. The specific targets examined for convergence were /p, o, i, s/ which are produced distinctly in each dialect. For instance, the duration of the vowel /i/ and the tongue height of word final /o/ vary across the two dialects. In the first study, segment duration, formant frequencies and MFCC were used to examine the convergence patterns. The results showed that speakers modified their own speech with exposure to dialectally distinct recordings. Specifically, MFCCs for /i, o, s/, formants for /o, i/ and duration for /o/ were good discriminators of the two dialects in these speakers. Considering that speech changes took place by exposure to a language variety in the absence of explicit repetition or interaction, the authors proposed that convergence is automatic. A second

study examined individual differences in speakers after convergence. Once speakers had converged to the speech of another dialect, they were asked to produce sentences themselves. Large individual differences were noted in convergence regardless of native dialect and dialect of exposure.

The adaptations lasted close to 10 minutes after the exposure, prompting the authors to suggest that convergence leads to durable changes in representations. The authors proposed that these findings indicate mimesis, not imitation, as the process responsible for PC. Mimesis results in durable changes to representations instead of instance-specific repetition that leads to imitation. It is interesting that the duration of /o/ demonstrated the strongest convergence even though it is not a dialectal marker. The authors proposed that the open or closed quality of /o/ (the actual dialectal marker) was less imitated than its duration because the tongue height in word-final /o/ is used as a sociolinguistic distinction between the two dialects. In other words, some dialectal markers may be more stable or more normalized in perception and thus less prone to convergence. This appears to contradict Goldinger's (1997) finding that lower frequency words are imitated to a greater extent. Perhaps, the difference in /o/ duration was more unexpected than the tongue height difference. Presumably the two dialects, although distinct, are not separated geographically, allowing for frequent exposure to both dialects. Other dialectal variations in the recordings may have cued the listeners to expect tongue height differences but the duration differences were novel and unexpected leading to greater convergence.

Another study that examined vowels during convergence found that men and women converged to 'model speakers' during vowel shadowing tasks (Babel, 2009a). The study asked subjects to repeat monosyllabic words that contained one of the vowels /i, æ, ɑ, o, u/ as produced by two model speakers who spoke the same dialect of

Californian English. This study tracked the Euclidean distance between the first and second formant frequencies at the vowel midpoint. Both male and female speakers converged more in the low, back /ɑ/ and / æ / vowels than in /i, o, u/. Both model speakers were noted for having the nasal split (/æ/ is more tense and has a lower F1 in nasal environments) though one model speaker's /ɑ/ had a very low F2. These factors lead the author to suggest that more word-specific production variants of these vowels were available to participants leading to a greater likelihood of imitation. Thus, in keeping with the automatic theory of accommodation, /ɑ/ and /æ/ were more likely to be imitated due to word-specific low frequency. Thus, using Goldinger's (1998) exemplar theory, these vowels were marked in the model speakers and created stronger episodes, which led to greater convergence.

VOT has also been examined as a target of convergence. Sancier and Fowler (1997) examined the speech of a female speaker of Brazilian Portuguese after a 4.5-month stay in the US, after a 2.5-month stay in Brazil and once again after 4 months in the US. Specifically, VOTs of the [t]/[t^h] and [p]/[p^h] variants were analyzed using recordings in English and Portuguese. Speakers of Brazilian Portuguese use only the unaspirated variety, [p] or [t] (*i.e.* extremely short VOT). This speaker demonstrated changes in VOT in the direction of the ambient language. Thus, her Portuguese VOTs drifted towards AE values when in the US and the VOTs of her English /p/ and /t/ drifted towards the Portuguese varieties when she spent time in Brazil. The authors suggested that this convergence of VOT values is due to imitation of gestures from the target language. The authors also raised a second possibility that the speaker detects the difference between L1 and L2 versions of each phoneme and uses the 'authentic' version most appropriate for the ambient language. However, these VOT drifts also take place in the direction of the dominant language of the environment. It would be interesting to test

whether these drifts are a result of articulatory efficiency considerations and would be different in different scenarios (*i.e.* conversing with a fellow Brazilian in the US or conversing with an American in Brazil). Finally, Nielsen (2008) examined imitation of artificially lengthened VOT in stops. Using a word-shadowing task, she showed that speakers were capable of imitating artificially lengthened VOT in stops. Moreover, participants were able to generalize the information from the token they were trained on, /p/, to a new phoneme, /k/, suggesting that convergence takes place at the word level as well as the feature level. Nielsen (2008) findings confirm Goldinger's (1998) prediction that imitation is a result of a generalization of word and feature level specifications.

Besides examining convergence patterns, studies such as the one by Sancier and Fowler (1997), described above, provide perceptual accounts for the underlying mechanism involved in PC. The authors suggest that this drift can be explained via the theories of direct realism (DR) and motor theory (MT) but not using a general auditory approach (GA). DR and MT posit gestures (or the intended neuromotor commands that lead to these gestures) as the invariant targets of speech whereas GA posits that speakers recover relevant acoustic information directly from the speech signal without mediation from invariants such as gestures. Thus, from a DR/MT perspective, these gestures (or the intended neuromotor commands) are the invariant targets for PC. The authors suggest that gestures are imitated in a manner similar to imitation of facial expressions. Once a perceiver sees a facial expression, it is possible to instruct one's own face to make that same expression. They also claim that GA is incapable of explaining the imitative process in terms of acoustic properties that map onto phonological categories. However, they do not elaborate why GA is incapable of explaining PC. Despite Sancier and Fowler's (1997) claim that GA is incapable of explaining PC, DR or MT are no more compatible with PC than GA. Explaining convergence using MT or DR must involve presuming that

there are invariant targets (*i.e*. gestures) that must be achieved during production with minimal noise. But positing underlying gestures is unnecessary to decode the spoken message from the speech signal (Diehl, Lotto & Holt, 2004). Thus, PC may be the result of something other than imitation of gestures. This prediction is supported by findings that both fine (feature-level) and gross (word-level) phonetic information is used during convergence, suggesting that imitation can take place beyond the level of phonemes (Nielson, 2008; Nenkova et al., 2008).

Given that requiring invariance is unnecessary for speech perception, and production and perception are directly linked (Chartrand, Maddux & Lakin, 2005), then speech is free to vary to suit task and speaker requirements predicting a purely imitative form of speech. Thus, convergence is the direct result of production following perception within the constraints of cognitive load, articulatory considerations and social factors. PC is a mimetic resultant product of the direct perception-production link in the face of biological, cognitive and social constraints while speakers are attempting to balance production and perception demands based on the communicative situation (Lindblom, 1990). In other words, we produce what we perceive constrained by our physiology and the demands of the task.

2.1.4.2 Suprasegmental correlates of PC

Convergence of suprasegmental features of speech has been studied significantly less than segmental adaptations. Ní Chiosáin (2007) reported mixed findings using two dialects of Irish (Northern and Southern). Using the synchronous speech paradigm (Cummins, 2003) in which speaker dyads must read a story simultaneously, she investigated how lexical stress placement, vowel duration and lenition of a specific class

of phonemes (the voiced nonpalatalized bilabial stops) changed in response to the task. Dyads were compared reading with speakers who matched their dialect and spoke the variant. Convergence was reported for vowel duration and stress but effects were small and exhibited large individual differences. For example, stress is always placed syllable initially in Northern Irish but is placed on the non-initial heavy syllable in Southern Irish. For words in which the second vowel is longer than the first, the two southern Irish speakers were perceived (by unspecified listeners) as placing word initial stress. For these speakers, the second longer vowel was also measured as shortened when compared to their intra-dialectal recordings.

Using the same paradigm, Krivokapic (2010) reported that speakers of British and American English converged in prosodic patterns of stress and intonation contours, noting that recently arrived British speakers converged more to American speakers than vice versa. Lewandowski (2012) used amplitude envelopes to measure phrase-level durational changes taking place to dyads' speech using native and non-native (L1 German) speakers of English. She found that speakers who approximated native-like pronunciation were more likely to converge with their English-speaking partners. This study is described further in Section 2.1.5.

Only two studies have examined convergence in rhythm (Krivokapic, 2013; Rao et al., 2011c). Using the synchronized speech paradigm, Krivokapic (2013) examined convergence of rhythm in four gender-matched dyads of Indian English (IE) and AE. Using a durational measure, which examined the duration of stressed syllables and feet to quantify rhythm, she found that one out of four pairs showed convergence. Specifically, one IE speakers' rhythm altered to the more AE style stress-timed pattern after the interaction. Rao et al (2011) examined rhythm convergence in linguistically variant groups using EMS + centroid, a spectral measure of rhythm. This study was the pilot for

the current study and examined one male and one female gender-matched pair for native language group with AE speakers, for dialectally variant group with AE and IE speakers and for mixed language group with L1 SP (L2 AE) and AE speakers. Analysis of this small dataset showed that the rhythm of the speakers in the native language group and the mixed language group converged whereas the rhythm of the speakers in the dialectally variant group diverged.

Although the preceding two sections discuss segmental and suprasegmental adaptations in speakers mostly from an automatic perspective, there is no reason why social considerations such as the need to align oneself with an interlocutor or to present oneself favorably to another would not explain the same findings. Social factors appear to be best suited in explaining divergence and are discussed below in section 2.1.5.

2.1.5 Social factors and phonetic divergence

Social considerations regarding convergence by speakers have often been noted via phonetic divergence, which has demonstrated across a variety of conditions such as dialectal and language variations and differences in attitudes and gender (Babel, 2009a, 2010; Bourhis & Giles, 1977; Kim et al. 2011; Namy, Nygaard & Sauerteig, 2002). The results suggest that divergence can be used to signal disagreement (Babel, 2010; Bourhis & Giles, 1977) as well as out-group membership. The same study by Babel (2009a) described above found that social factors, such as attitudes towards race, affect convergence. In order to study the effect of race on PC, the target words used in the shadowing task were produced by two model speakers, one African-American and the other Caucasian. Female speakers who were rated as 'pro-black' (via a race bias test called implicit association task or IAT) converged more to the African-American

speaker's speech than to the Caucasian-American's speech. The fact that convergence in this study was phonetically and socially selective suggests that it is not an entirely automatic process. Instead, she posits an automatic theory of convergence where implicit social factors apply at an unintentional level.

Similar socially motivated findings were also reported by Bourhis and Giles (1977) and Babel (2010). The first study noted that people who disagreed with their interlocutor showed divergence. Bourhis and Giles (1977) studied the speech of Welsh adults who were either attending Welsh language and Welsh culture classes, or just Welsh language classes as they interacted with an English speaker. For some participants, the English speaker was presented as someone who questioned the function of Welsh in the present day. The group taking both the Welsh language and culture classes diverged from the English speaker, whereas the group that was taking only Welsh language classes converged with the English speaker. Thus, disagreement with the English speaker's point of view on Welsh caused the participants in the language and culture class group to diverge. Babel (2010) reproduced this study using vowels that are distinct in New Zealand English (NZE) and Australian English (AuE) (in words such as KIT and TRAP respectively). Participants from New Zealand who were rated as being pro-Australian (via the IAT) were more likely to converge to Australian speech. In fact, a participant's attitude towards Australians was the only significant predictor of convergence. This was found despite the fact that NZE speakers were sometimes told that the AuE speaker had insulted New Zealand.

The studies described above show that social factors like a speaker's attitude contribute to whether they converge with an interlocutor. Attitudes towards race, however, are not the only relevant social factors. Dialectal and language differences have also been found to affect PC (Lewandowski, 2012; Babel 2010; Krivokapic, 2013). For

instance, a study using diapix by Kim et al. (2011) examined convergence across languages and regional dialects as a function of interlocutor linguistic distance. The authors defined interlocutor linguistic distance between interlocutors as the following:

- Close: both speakers share L1 and speak the same regional dialect
- Intermediate: both speakers share L1 but speak differing regional dialects
- Far: speakers do not share L1

Speakers in the close and intermediate conditions consisted of two pairs of AE speakers and two that spoke Korean. Speakers in the far condition consisted of AE speakers talking to native Korean and Chinese speakers. Convergence was evaluated using XAB tasks with a separate set of listeners. Results suggest that speakers were more likely to converge if they belonged to the close condition. In other words, speakers that shared the same language and the same dialect were more likely to converge than speakers who spoke the same L1 but not the same dialect and those who spoke different L1s. The authors speculated that the need for intelligibility combined with the increased processing load due to non-native speech production and perception would lead to the inhibition of convergence in the intermediate and far groups.

Lewandowski (2012) examined the role of phonetic or pronunciation talent in PC by comparing the speech of one native model speaker each of BE and AE with non-native speakers of English who were native speakers of German. Proficiency tests were conducted as part of another study but no other details are provided regarding the evaluation. She used two comparisons for speaker pairs: early and late speech tokens from the dialogue during a diapix task, and pre and post task tokens that contained words from the diapix task. Using the amplitude envelope, which tracked amplitude fluctuation over the course of the word pronunciation, she found that non-native speakers of English who were more talented at pronunciation converged with their native English-speaking

partners whereas less talented non-native speakers either diverged or showed maintenance. This convergence was only noted in early and late speech from the dialog; pre-task and post-task data did not reveal any indication of convergence.

Sex-based differences have also been noted in PC (Babel, 2009a; Babel, 2010; Namy et al., 2002). Babel's study (2010, described above in Section 2.1.7) used speakers of NZE and AuE to examine the effect of varying attitudes towards their interlocutor on convergence. This study used the same methodology as that of Babel (2009a). She found that women in general converged whereas men showed vowel selectivity. They converged on vowels in words such as DRESS, BARN, STRUT and THOUGHT but diverged for KIT and TRAP. The vowels in KIT and TRAP are noted as being salient distinctions between NZE and AuE. Namy et al. (2002) studied gender differences in speakers and listeners during word-shadowing tasks. They found that females converged more than males did when shadowing both male and female word tokens presented over earphones. In addition, female speakers converged to a greater degree to male tokens than to female tokens. Moreover, female listeners were more adept at detecting accommodation than male listeners. The authors speculated that this may be due, in part, to differences in perceptual sensitivity and attention to indexical features in the speakers. In contrast, two studies that use map tasks by Pardo and colleagues (Pardo, 2006; Pardo, Jay, & Krauss, 2010) have found that male dyads converged more than female dyads. Pardo (2006) used AXB tasks to detect PC while Pardo et al., (2010) used acoustic analyses of articulation rates and vowel formant euclidean distances.

Perceptual tasks have shown that PC is also subject to differences due to role (whether the speaker is providing instructions or receiving and reproducing instructions). Pardo (2006) showed that the speech of female receivers did not converge to female providers but that of male receivers did converge to their male partners. Pardo et al.

(2010) extended these findings by showing that asking speakers to explicitly imitate resulted in varying results based on role. If providers were explicitly asked to imitate the speech of their partners, female dyads did not converge whereas male dyads did. However, both male and female dyads showed convergence if receivers were asked to imitate speech. The authors call for further research in order to determine the reasons for this pattern of role-based differences in convergence. Both the aforementioned studies by Pardo and her colleagues have used map tasks to study convergence. However, role-based differences have also been noted in the absence of explicit role assignment. Kim et al. (2011), who used a diapix task to study the effect of linguistic distance on convergence, noted role-based differences in their findings. They speculated that the speakers may have set up spontaneous 'leader' and 'follower' roles that led to these differences.

2.2 VOWELS

In this section, past research on vowels is described briefly followed by analyses used in notable PC studies. Vowel quality is defined by a number of properties such as vowel height, front/back distinction, and nasalization, among other properties. The height of a vowel is dependent on the placement of the tongue with respect to the palate or the jaw. Thus the /a/ vowel (as in fa**ç**ade) is lower than the /i/ vowel (as in f**ee**l) because it is produced by placing the tongue lower than in /i/. In other words, /a/ is produced with a more open set of articulators than /i/. For this reason, height is also referred to as vowel openness. Similarly, backness of a vowel is dependent on the position of the tongue in the mouth during articulation. For example, /i/, a front vowel, is produced with the tongue body towards the front of the mouth whereas /u/ (as in f**oo**l), which is a back vowel, is

produced with the tongue placed further back in the mouth than /i/. Nasalization in vowels is created by leaving the velum open so that air may pass through the nasal cavity during articulation.

Research on vowel identification and classification as followed one of two approaches: static and dynamic both of which as described below in Sections 2.2.1 and 2.2.2

2.2.1 Static measures of vowels

The static approach demonstrates that vowel quality and formant values are correlated (Peterson & Barney, 1952; Hillenbrand, Getty, Clark & Wheeler, 1995). Specifically, two features of vowel quality, height and backness, are correlated with the second and first formants respectively. Increasing F1 leads to a decrease in vowel height and an increasing F2 leads to a vowel that is more fronted in quality (Figure 2.1). These measurements were taken once the vowels had reached a steady-state condition but measurements are also taken at the onset of the vowel, center of the vowel or the offset of the vowel. Hillenbrand et al. (1995) cautioned that since vowels are subject to long-term change, these values can only describe vowels within a dialect at a particular point in time of the dialect's history. They cannot be considered canonical vowels for a given language.

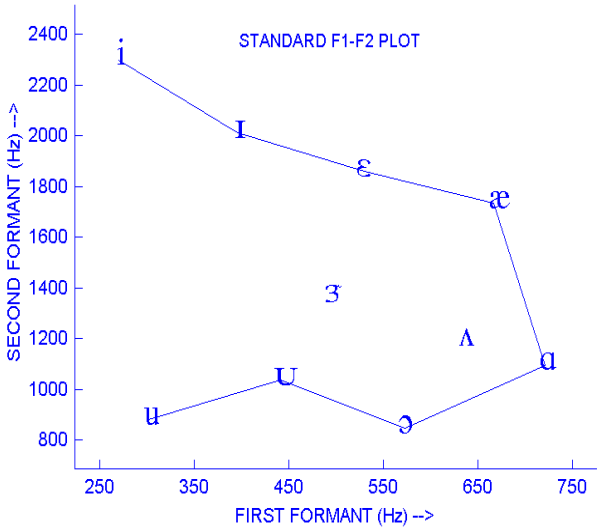


Figure 2.1: Vowel averages of first and second formant values (from Hillenbrand et al., 1995)

Furthermore, Figure 2.2 shows that despite the correlation between formant values and vowel quality, vowels do not fall into neatly separated areas on the formant chart. Instead, there is large overlap between vowel categories. Hillenbrand et al. (1995) noted that adding spectral and durational information is helpful in further separating these categories. They also noted that spectral information is better at the task than durational information alone.

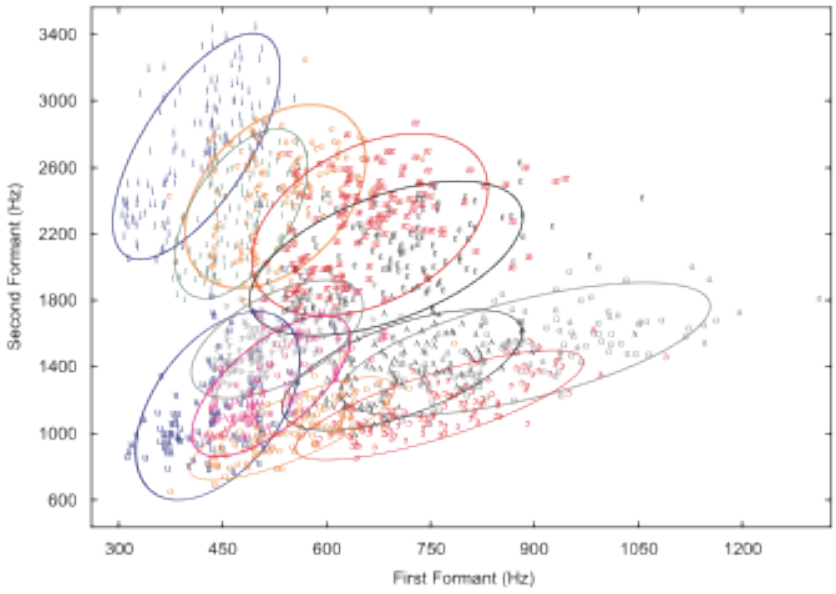


Figure 2.2: Formant frequencies for vowels (from Hillenbrand et al., 1995)

2.2.2 Dynamic measures of vowels

Improvement in identification from the addition of durational or spectral information underlines the insufficiency of static methods of measuring vowel formants, be that at the steady state, midpoint or onset (Hillenbrand, 2011). An alternate approach to vowel classification is capable of incorporating dynamic information into the analysis. The most common spectral approach to vowel categorization is called the vowel internal spectral change or VISC (Nearey & Assmann, 1986; Morrison & Nearey, 2007, Hay et al., 2006). VISC is capable of measuring slow, durational changes to formant frequencies of vowels. It is calculated using one of the following three methods:

1. Onset + offset: the difference between the vowel onset and vowel offset values of the first formant. This measure, also called dual target, presupposes that only the end points of a vowel are relevant to vowel discrimination.

2. Onset + slope: the difference between the onset and the steady state value of a vowel's first formant over time. This measure accounts for the rate of change in a vowel.
3. Onset + direction: the direction of the first formant movement. Diverging, converging or flat are represented as -1, 1 or 0. This measure takes only the general direction of change of vowel formants into account.

Nearey and Assmann (1986) used perceptual listening tasks and pattern recognition models to test the role of VISC in vowel discrimination. Using the vowels /i/, /ɛ/, /æ/, /e/ and /o/ in an isolated context, they found that all three methods were able to characterize listeners' ability to distinguish vowels. More specifically, the onset + offset and onset + direction methods were marginally better than the onset + slope approach. However, more recent research by Morrison and Nearey (2007) comparing the three approaches found different results. Using the vowels /e/, /i/ and /ɛ/ in a /bVpə/ environment, they tested the three methods of calculating VISC using synthetic stimuli in a perceptual listening task. Their study revealed that onset + offset is the most useful for listeners when distinguishing between vowels.

2.3 RHYTHM

2.3.1 Origin and history

Initial research on rhythm focused on the idea that prosodic rhythm consists of isochronous units of speech. Pike (1946) and Abercrombie (1965) were among the first to attempt to define linguistic rhythm based on isochrony of inter-stress intervals or syllables. Pike (1946) described English as a stress-timed language and Spanish as a syllable-timed language where stress-timed languages have isochronous inter-stress

intervals and syllable-stressed languages have isochronous stressed syllables. Abercrombie (1965) attempted to describe this isochrony in physiological terms based on the rhythm of the muscles involved in breathing. In these terms, rhythm is the succession and coordination of stress and chest pulses. Syllable-timed languages demonstrate equally spaced chest pulses or isochronous syllables. Stress-timed languages on the other hand have equally spaced stress pulses or inter-stress intervals. Bloch (1950) introduced a third classification for languages such as Japanese and Tamil: mora-timed. In these languages, morae, which are sub-units of syllables that consist of a preceding consonant followed by a vowel, were thought to be isochronous.

Subsequent attempts at isolating isochronous units of rhythm in the acoustic signal were largely unsuccessful (Dauer, 1983; Lehiste, 1977). Inter-stress intervals are just as likely to be isochronous in syllable-timed languages as in stress-timed languages. The idea of an 'objective isochrony' was altered to that of 'subjective isochrony' (Lehiste, 1977; Dauer, 1983). Lehiste (1977) suggested that humans are predisposed to impose a rhythmic structure on sequences of inter-stress intervals. Hence, isochrony may be a perceptual phenomenon that is guided/determined by a listener. Inter-stress intervals (and syllables) do not show any principled patterning in stress-timed languages or syllable-timed ones making classifying languages as acoustically isochronous impossible (Dauer, 1983). Instead, rhythm may be a consequence of linguistic circumstance (Dauer, 1983; Lehiste, 1977; Krull & Engstrand, 2003). Dauer (1983) pinpointed three factors that can lead to rhythmicity: varied syllable structure (presence or absence of complex consonant clusters) and the influence of stress on vowel duration (stressed vowels are longer than unstressed vowels). She also proposed that instead of being part of discrete rhythm classes, languages can be placed along a rhythmic continuum where syllable-timed and stress-timed languages lie at opposite ends. This account of rhythm allows for

intermediate languages that may be more or less stress- (or syllable-) timed such as Polish, which allows consonant clusters but not vowel reduction.

2.3.2 Traditional duration measures of rhythm

Cross-language investigations have attempted to relate these phonological properties of varied rhythmic types to speech signal properties in a phonetic description (Ramus, Nespor & Mehler, 1999; Low et al., 2000; Grabe & Low, 2002). Ramus et al. (1999) hypothesized that differences in linguistic structure must lead to rhythmic differences in languages. For example, English has a large number of syllable types and demonstrates vowel reduction. On the other hand, Spanish has fewer syllable types and does not allow vowel reduction. To test this hypothesis, Ramus et al. (1999) examined eight languages: English, Dutch, Polish, French, Catalan, Spanish, Italian and Japanese. Five short declarative sentences, matched for syllable count and average duration, were spoken by four female native speakers of each language. Assuming that stress-timed languages have greater contrast in vowel duration between stressed and unstressed syllables and that stress-timed languages have greater variation in the complexity of consonant clusters (or duration of consonant intervals), four variables were proposed and analyzed for their ability to mark rhythm:

- %V: proportion of vocalic intervals within a sentence
- %C: proportion of consonantal intervals within a sentence
- ΔV: the standard deviation of the duration of vocalic intervals in each sentence
- ΔC: standard deviation of duration of consonantal intervals in each sentence

The descriptive plot of %V vs. ΔC in Figure 2.3 demonstrates how language types are classified. Stress-timed languages (*e.g.* English) tend to have greater vowel reduction

(lower %V) and consonant interval variability (larger ΔC). Conversely, syllable-timed languages do not allow vowel reduction (higher %V) and less complex syllable structure (lower ΔC). The results showed that %V vs. ΔC was able to separate the eight languages into three groups (Figure 2.3). English, Dutch and Polish fell into one group; Spanish, Italian, French and Catalan formed a second group and Japanese formed the last group. These three groups align with the stress-, syllable- and mora- timed groups described by previous research.

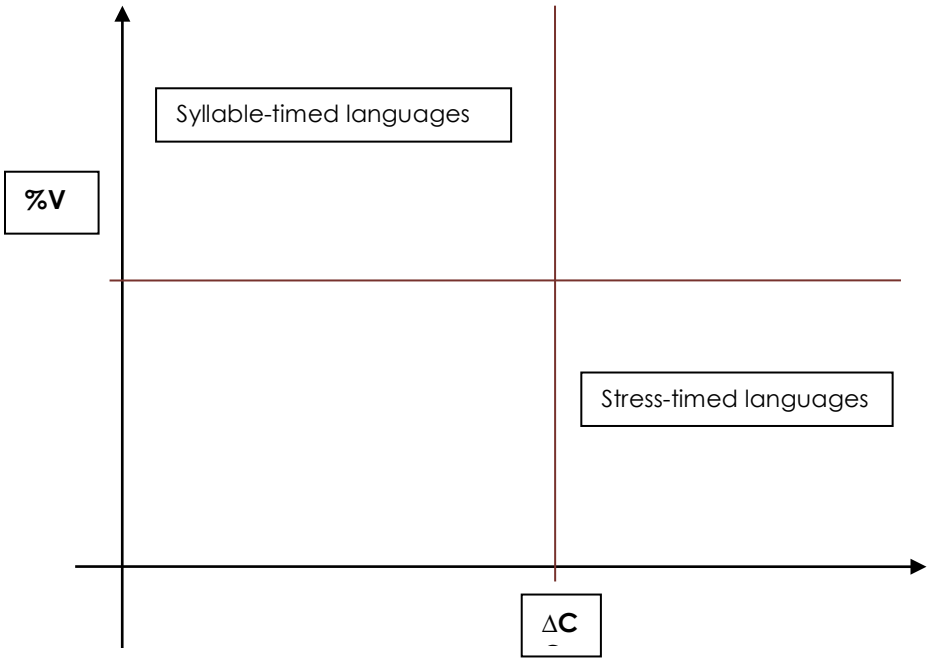


Figure 2.3: Language classification based on rhythm as measured by %V and ΔC

The above measure however did not account for inter-speaker speaking rate differences. To address this, Dellwo (2006) implemented a measure that incorporates

speech rate with syllable internal measures of rhythm. The variation coefficient for ΔC, varcoC, incorporates rate of speech information with Ramus et al.'s (1999) ΔC metric. VarcoC is defined as the percentage of standard deviation of consonantal intervals per average consonantal interval duration and is calculated using the formula: ΔC/mean duration of consonantal intervals. Using this formula, English, German and French sentences were produced at varying speech rates and analyzed as slowest, slow, normal, fast and fastest. Results were plotted along the %V vs. varcoC dimensions. Although varcoC was able to separate English and German sentences based on the differing rates of speech, it was unable to do so with the French sentences. This suggests that varcoC scores for English and German are affected by changes in speech rate whereas it remains relatively stable for French across varying speech rates.

%V, ΔC, ΔV, varcoC and varcoV only consider vowel and consonant variations within each syllable. They are considered syllable internal acoustic measures of rhythm. In order to obtain a measure of the inter-segmental variability across phrases, the pairwise variability index (PVI) was proposed by Low, Grabe and Nolan (2000) and subsequently Grabe and Low (2003). PVI is defined as the variability in durations of successive vocalic and intervocalic segments. PVI measures eliminate the possibility of spurious variability being introduced from changes in speaking rates and between-speaker differences that are present in %V, ΔC and ΔV. Grabe and Low (2003) examined both raw PVI (rPVI) and PVI normalized for speech rate (nPVI) in 18 languages, which included stress-timed, syllable-timed, mixed and unclassified languages. Vocalic nPVI plotted against intervocalic rPVI were capable of separating languages based on classic rhythm categories (Figure 2.4). Notably, mixed languages showed some interesting patterning. For example, Polish has a vocalic nPVI very similar to French (syllable-timed) but an intervocalic rPVI value that is very different from it. Similar patterning was

noted for Catalan, another mixed language. Taking this overlap between languages into account led the authors to support Dauer's (1983) idea of a continuous rhythm dimension. Vocalic nPVI combined with %V may provide a better measure of rhythm because it combines overall vowel time with vocalic variability.

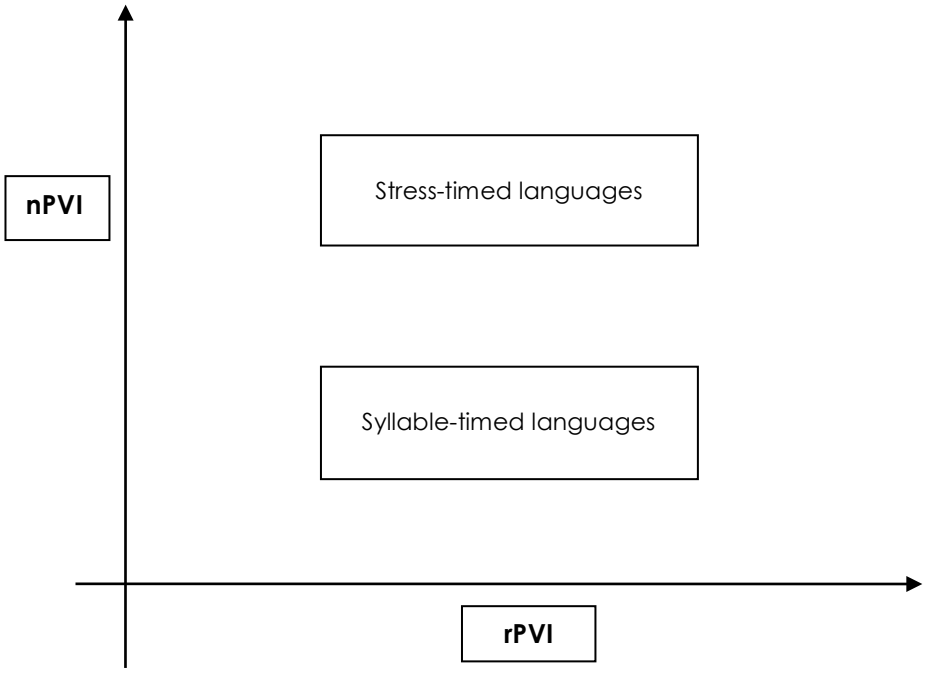


Figure 2.4: Language classification based on rhythm as measured by nPVI and rPVI

Since the above-described metrics are quantifications of consonantal and vocalic variability, it is important to consider factors that can alter this variability. Research reveals that elicitation style, speaking style and even the person taking acoustic measurements can affect rhythm metrics (Wiget et al., 2010; Smiljanic & Bradlow, 2008; Krull & Engstrand, 2003). Wiget et al. (2010) evaluated %V, varcoV, and nPVI for the effect of sources of variation due to individual speaker differences, recorded sentence

materials and measurement takers. They reported that all three might contribute to variability in rhythmicity with recorded sentences themselves adding the greatest source of variability in rhythm metrics. Speaking style also affects prosodic rhythm. Though speaking style (speaking clearly vs. conversationally) may affect durational properties of speech (Krull and Engstrand, 2003), Smiljanic and Bradlow (2008) showed that rhythm measures such as %V, %C, varcoC and varcoV remained unchanged as speakers changed their consonant and vowel intervals to the same degree (though Rao & Smiljanic, 2011b showed that speaking style does alter spectral measures of rhythm).

2.3.2.1 Rhythm variation across linguistic background

Besides typological studies, rhythmic differences have also been found in dialectal variants and non-native speech (Low, Grabe & Nolan, 2000; White & Mattys, 2007). Low et al. (2000) used the PVI metrics to examine rhythmic differences in two dialects of English that were thought to share rhythm characteristics: Singaporean English (SE) and British English (BE). They found that, as with other syllable-timed languages, SE demonstrates lower PVI values. This suggests that SE has less vowel reduction and consonant variability than BE which has higher PVI values. Additional research using PVI demonstrated dialectal differences in rhythm within Native American varieties of English and standard American English (Coggshall, 2009), standard American English and Hispanic English (Carter, 2005), Brazilian Portuguese and European Portuguese (Frota & Vigário, 2001) and two Peruvian dialects of Spanish (O'Rourke, 2008).

Differences in the rhythm of native (L1) and non-native (L2) speakers of languages have also been explored (White & Mattys (2007); Carter, 2005). These studies

found that speakers' L1 rhythm can affect their L2 rhythm. White and Mattys (2007) analyzed the rhythm of L1 speakers and competent but accented L2 speakers of languages that fall in varying rhythm classes with the expectation that L1 rhythm would influence L2 rhythm. Using Dutch, English and Spanish as the languages of comparison, they found that L2 speakers display rhythm metric scores that are intermediate to their L1 and the target language but only if the two languages differ rhythmically. L2 speakers of a rhythmically similar language do not make the small changes needed for the L2 language. The above findings are supported by Carter (2005) who compared the rhythm of L1 speakers of English and Spanish and L2 speakers of English who were L1 speakers of Spanish. Using PVI, he found that L2 English scores are intermediate to L1 Spanish and L1 English rhythm scores. Thus, non-native speakers of a language display intermediate rhythm between their native L1 and target L2 properties when the two languages are rhythmically distinct.

2.3.3 Issues with traditional measures of rhythm

The research outlined above suggests that %V in combination with nPVI can be informative in classifying languages and L1/L2 rhythm. However, there are disadvantages to using these measures to classify rhythmic categories. Arvaniti (2009) argues that current rhythm metrics are insufficient because they are simply a crude measure of timing. She states that rhythm and timing are often confounded and attempts at quantification only consider consonant and vocalic variability. Although these metrics have some success in separating prototypical languages, they often fail when non-prototypical languages are included. Even prototypical languages do not always pattern in the expected manner. For example, %V scores for English and Spanish are not

significantly different. On the other hand, nPVI fails to separate German from Korean or Greek even though these languages are perceptually distinct. Patel (2008) supports the view that rhythm must be separated from periodicity. He points out that all rhythmic patterns are not periodic, although all periodic patterns are rhythmic. An example of a rhythmic pattern that is not periodic is Ghanaian drumming which features a repeating but non-isochronous pattern. Instead he defines rhythm as the systematic patterning of sound in terms of timing, accent and grouping (2008; 96). In order to obtain a complete picture of linguistic rhythm, measurements must go beyond relative syllable strength to include groupings and patterns of prominence, such as syllable distribution (Arvaniti, 2009; White & Mattys, 2007).

2.3.4 Psychological reality of rhythm

Although acoustic correlates of rhythm have been elusive, perceptual studies such as Ramus and Mehler (1999) and Ramus et al. (2003) have demonstrated the psychological reality of rhythmic differences in languages. They found that French subjects were able to discriminate between English and Japanese sentences based on rhythm alone. Participants were asked to distinguish between sentences of two artificial languages Sahatu and Moltec. Using resynthesized sentences, participants were placed in one of four separate conditions in which sentences were modified in one of four ways:

1. *saltanaj*: global intonation, syllabic rhythm and broad phonotactics were preserved but all nonprosodic lexical and syntactic information and specific phonetic and phonotactic information was deleted by replacing all fricatives with /s/, all stops with /t/, all liquids with /l/, all nasals with /n/, all glides with /j/, and all vowels with /a/.

2. *sasasa*: syllabic rhythm and intonation information were preserved but all other information was deleted by replacing all consonants with /s/ and all vowels were replaced with /a/.
3. *aaaaa*: the intonation information alone was preserved by replacing all phonemes with /a/.
4. *Flat sasasa*: the fundamental frequency was held constant preserving only the syllabic rhythm but deleting any intonation information available in the sasasa condition. All consonants were replaced with /s/ and all vowels were replaced with /a/. In addition, all sentences were also resynthesized with a constant fundamental frequency of 230 Hz.

French listeners were able to distinguish sentences from the two languages in all conditions except *aaaaa*: the third condition in which only the intonation information was retained. However, native English speakers were able to distinguish between sentences in this condition (as well as the rest) suggesting that isolated global intonation information is only useful if the speaker has knowledge of one of the languages to be distinguished (albeit, unconscious because listeners were not told these were English and Japanese).

Ramus et al., 2003 used the flat sasasa version of speech from the above experiment to test whether speakers could tell languages apart that shared rhythmic properties to varying extents of their L1. Using native speakers of French they tested discrimination between 7 language pairs: English-Spanish, English-Dutch, Polish-English, Polish-Spanish, Catalan-English, Catalan-Spanish and Polish-Catalan. They found that listeners were able to discriminate between all language pairs except between Spanish-Catalan and English-Dutch. The two pairings that were not discriminable belong to the same rhythm class suggesting that rhythmic differences were discriminable across languages. Discrimination between English-Spanish and English-Polish and English-

Catalan is of particular relevance to this study. English and Spanish are the two languages being used in the current study based on their rhythmic distinctions. IE, which is also used in this study, has been classified as rhythmically mixed like Polish and Catalan. Since listeners were able to tell English apart from other rhythmically mixed languages, they should be able do so with IE as well.

2.3.5 Spectral measures of rhythm

In an attempt to move beyond segmental interval duration measurements and their shortcomings, Tilsen and Johnson (2008) employed a spectral analysis of rhythm for conversational and citation style speech from the Buckeye Corpus (Pitt et al., 2005 as cited in Tilsen & Johnson, 2008). This procedure low-pass filters speech with a cutoff of 10 Hz, calculates the amplitude of this filtered signal and calculates the Fourier transform of this amplitude envelope. The authors call this the power spectrum of the amplitude or the rhythm spectrum. Note that this is the power spectrum of the amplitude envelope of the filtered speech signal, which contains only low-level amplitude fluctuation information but lacks any phonological and indexical information contained in the power spectrum of the unaltered signal as used by Lewandowski (2012). Tilsen and Johnson (2008) plotted the frequency and amplitude of the most prominent peak of the rhythm spectrum of speech tokens, approximately 400 ms in duration. The authors found that the 2-4 Hz range within 40-50 dB range of amplitude was most informative for English. Comparisons of citation and conversational forms of sentences showed a positive relationship between vowel and consonant deletion and rhythmicity. Conversational speech, which tends to be more rhythmically variable, demonstrated greater vowel and consonant deletions and had a concentration of energy in the 3-4 Hz band of the rhythm

spectrum. Citation style speech, which tends to be less rhythmically variable, was associated with a lack of consonant and vowel deletions denoted by the presence of energy in the 2-3 Hz region of the rhythm spectrum.

Measures of linguistic rhythm such as ΔV, ΔC, PVI and the variation coefficients are labor intensive and make linguistic presumptions about the language being examined. More importantly, these metrics ignore non-linguistic information such as pauses and other verbal hesitations, which are critical in distinguishing dysarthric from non-dysarthric speech (LeGendre et al., 2009). Such shortcomings of traditional measures of rhythm prompted Liss et al. (2010) to propose the use of the envelope modulation spectrum (EMS) in distinguishing between clinical and non-clinical speech. EMS describes the low level modulations of the speech signal by decomposing the periodicity of the amplitude envelope of the signal into its component frequencies. The rhythm spectrum approach described above is similar to EMS with two key differences. Firstly, EMS used a 30 Hz cut-off for the initial low-pass filter whereas the rhythm spectrum uses a 10 Hz cut-off. Secondly, EMS was calculated in speech filtered at octave bands up to 8000 Hz whereas the rhythm spectrum was only calculated in the original speech signal. EMS has been used with a discriminant factor analysis (DFA) to identify the most informative features for rhythm (Liss et al., 2010; LeGendre et al., 2009). EMS was successful in distinguishing between dysarthric and control speech, different types of dysarthric speech, individual speakers and sex of speakers of nonclinical speech (LeGendre et al., 2009).

In a preliminary study using a spectral measure of rhythm, Rao and Smiljanic (2011) demonstrated that EMS is capable of distinguishing between languages and speaking styles. They used a weighted mean of the power spectrum (henceforth, EMS + centroid) to analyze the rhythmic characteristics of two typologically distinct languages:

English (stress-timed) and Croatian (mixed type). Using sentences spoken in clear and conversational speaking styles, they found that the EMS + centroid measure was able to separate speakers based on language and speaking style. EMS + centroid is used to analyze rhythm in the current study; it will be described in greater detail in Section 3.4.2 of the methodology chapter.

2.3.6 Advantages of spectral measures

Spectral approaches to rhythm analysis are superior to those based on segmental intervals for various reasons. The need for phonology-dependent criteria for identifying syllables or phrases (or even consonants and vowels) is eliminated as are interval measurements, which can be difficult to identify in running speech and time consuming to extract. Interval measures also exclude non-phonetic information such as pauses and dysfluencies, which can affect rhythm. Spectral analyses, on the other hand, take the entire signal into account and ensure the inclusion of all relevant data. Thus, these approaches account for both relative syllable prominence and syllable structure (*e.g.* stress placement, intonation differences, cross-dialectal differences in vowel reductions etc.). Moreover, as Tilsen and Johnson (2008) point out, the spectral approach has the additional benefit of describing rhythm in terms of frequency prominence and representing it at the phrasal and utterance-level time-scales.

2.4 CURRENT STUDY

PC has been demonstrated in cross-linguistic comparisons (Kim et al., 2011; Lewandowski 2012, Babel 2010). However there is a limited amount of data on L1-L2 interactions and also in rhythm adaptations that occur during interactions. Furthermore,

there is a lack of information on global changes taking place to a dyad's vowel set due to an interaction. While vowel specific changes have been noted in dyads, no study has compared speakers' vowel sets. Is it the case that their vowel systems as a whole change due to the interaction or do they remain unchanged? The current study addresses these deficiencies in PC research by examining speech adaptations in speakers of AE, IE and SP using novel measures of vowel and rhythm convergence both of which measure gross, systemic changes taking place to a speaker's vowels and rhythm.

There are four features of the current study that benefit from spectral approaches to rhythm:

1. A small sample size (12 speakers of each sex in each language condition). Wiget et al. (2010) showed that there is a lack of support for interval measures for small sample sizes and longitudinal changes.
2. Mixture of conversational and citation speech. Speaking style and the dataset itself may be a source of rhythmic variability (Wiget et al., 2010; Rao & Smiljanic, 2011; Krull and Engstrand, 2003).
3. Short-term longitudinal changes over a time course of approximately 40-60 minutes.
4. The languages used in the current study: American English (AE), Spanish and Indian English (IE). Classical measures of rhythm are capable of reliably separating English and Spanish. However, with the exception of Tamil, little research exists on Indian languages. Patel (2008) found that Hindi is syllable-timed and Pingali (2010) suggests that IE is neither syllable- nor stress-timed (*c.f.* Fuchs, 2012). As described above, unclassified or mixed languages such as IE are precisely where segmental analyses tend to fail.

Therefore, the inclusion of conversational speech, small sample size in each language condition and a rhythmically mixed dialect of English make a spectral approach to rhythm analysis more appropriate for the current study.

Chapter 3: Methodology

3.1 PARTICIPANTS

This study tested convergence in speech that is native, non-native and dialectally distinct. The three language varieties were selected because they are all rhythmically distinct. Spanish has been classified as syllable-timed whereas IE has both syllable- and stress-timed characteristics (Ramus et al., 1999; Pingali, 2009). Interactions between native speakers of AE and Spanish will yield information on convergence between native and non-native speakers of differing languages whereas interactions between native speakers of AE and native speakers of IE will yield information on convergence between speakers of differing dialects of the same language with the focus on vowels and rhythm.

72 speakers total (36 male and 36 female), between the ages of 18-60 years, were recruited from the University of Texas - Austin (UT) campus and the Austin metropolitan area. Participants did not have any known speech or hearing impairments at the time of recording, however one female AE speaker identified as dyslexic. They belonged to one of the following groups: (a) a native monolingual speaker of AE (NS_{AE}), (b) a native speaker of Spanish and non-native speaker of English (NN_{SP}) or (c) a native speaker of Indian English (NS_{IE}). Participants were recruited using flyers, emails and word of mouth in accordance with IRB protocols. All participants were unfamiliar with each other at the time of the experiment.

Due to the geographical location of recruitment, native speakers were expected to have a general AE accent with some southern features but AE dialect was not explicitly controlled. All non-native speakers were asked to self-rate their proficiency in English on a Likert scale of 1-7 via email prior to their arrival for the experiment. Flege, Mackay and Piske (2002) confirmed the efficacy of self-reported scores for language dominance and the ability of a speaker to assess their own proficiency of a language. Only those Spanish

and IE speakers who rated their ability to speak and understand AE in the range 3-5 were invited to participate in this experiment. As a result, low proficiency speakers who may not have the sufficient experience with the target language to converge with native speakers were excluded (Kim et al., 2011). As part of the prescreening process, speakers were asked four questions based on the LEAP-Q (Language Experience and Proficiency Questionnaire) questionnaire (Marian, Blumenfeld & Kaushanskaya, 2007):

1. On a scale of 1-7 (where 1 is none and 7 is perfect), how would you rate your ability to speak English?
2. On a scale of 1-7 (where 1 is none and 7 is perfect), how would you rate your ability to understand English?
3. On a scale of 1-7 (where 1 is slightly or none at all and 7 is completely different), in your perception, how different is your accent from the American accent?
4. On a scale of 1-7 (where 1 is not very often and 7 is everyday) how frequently do others ask you to repeat something you said based on your accent?

A more detailed demographic questionnaire was administered after obtaining consent.

3.2 RECORDING APPARATUS

Each participant was fitted with a head-mounted Shure SM10A unidirectional microphone. Recordings were made directly to a Dell PC computer using a MOTU Ultralite MK3 digital/analog convertor with Adobe Audition on separate channels for each speaker. Recordings took place in a sound attenuated booth at a sampling rate of 44.1 kHz with 16-bit resolution. Elicitations for recordings and map completion tasks were presented using Microsoft PowerPoint.

3.3 MATERIALS

3.3.1 Language background and demographic questionnaire

A questionnaire to obtain information about their language(s) usage and dominance was administered to the participants when they arrived for the experiment. This questionnaire had been modified from the LEAP-Q (Marian et al., 2007, see Appendix C). It was used to gather information about the linguistic backgrounds of the speakers such as dialect of primary language, knowledge of second or third languages, language dominance, and frequency of usage, among other questions.

3.3.2 Vowels

Participants were asked to produce six vowels that sample the AE vowel space. They were presented in an hVd context within a carrier sentence: "*Say the word* ___ *again*". Each vowel was elicited five times in random order. The words used for this study are listed in Table 3.1 below. The elicited vowel is listed in the adjacent column in the International Phonetic Alphabet (IPA). Participants read this randomized list before and after the interactive map task to provide pre-task and post-task data for vowel analysis. If non-native speakers were unfamiliar with a word, they were given a chance to practice it before it was recorded (if possible, the meaning of the word was also provided).

Word	Vowel (IPA)
Had	/æ/
Hod	/ɑ/
Heed	/i/
Who'd	/u/
Hayed	/e/
Hoed	/o/

Table 3.1: List of vowels in an hVd context

3.3.3 Recording paragraph

In addition to the vowel list and map tasks, participants were also recorded reading a short paragraph. This 'recording paragraph' contained a short story that incorporates landmark items from the below-described map tasks. Participants read this story as part of the pre-task and post-task recordings. Landmark items were presented in paragraph form to record each subject's initial and final rhythm in a story format that allowed for speech that is closer to spontaneous connected speech. By using landmark items from the maps in the recording paragraph, speakers were provided opportunity to repeat these phrases creating stronger convergence. Analysis of the paragraph's rhythmic characteristics would detect any rhythm changes that may occur over the course of the map task. As a result, rhythm characteristics were consistent across pre-task, during-task and post-task repetitions. The recording paragraph is shown below (the landmark items that also appear on the maps are shown in bold):

> One day, Sara put on her **cowboy boots** and went for a walk. She left her **sleeping cat** at home. She walked over the **suspension bridge** and came

to a **white mountain**. This mountain is the **highest viewpoint** in the region. She saw some **spotted deer** at the base. There were also some **bumble bees** around wildflowers. She was surprised to see a **black bat** dart out of the **forest** and disappear quickly into the **old temple**. There used to be a **diamond mine** near the mountain but it is now abandoned. On the way back, Sara crossed the **pedestrian bridge**, which is actually an i**ron bridge** to stop by the **large lake**. Back at home, she stopped to check the **banana tree**. The bananas were still unripe so she decided to wait to eat them.

3.3.4 Maps

Four map pairs based on those developed by Anderson et al. (1991) were used in this study. These are included in Appendix B. Each map features a route around landmark items (*e.g.* mountains) that is present in one version of the map but absent from the other. Each landmark item has a descriptive label, either a word or a phrase, under the image, to elicit the production of the target words. The map tasks used in this study have been modified from the Anderson et al. (1991) version to include items that are familiar to non-American speakers. Recordings from these map tasks provided data for the during-task analysis of rhythm.

In order to ensure familiarity with the new items, seven native, monolingual speakers and seven non-native speakers of AE were asked to examine each image on the map. They were asked to note if they were unfamiliar with the image or word or phrase that was listed under it. If the issue was the word or phrase being used, an alternate suggestion was requested from the participant. Moreover, they were also asked to provide

any feedback about each map as a whole. If more than one speaker noted an issue with an image or word, it was altered to the word or image described by the speakers. The finalized maps were also used in practice sessions by a pair of NS_{AE}- NS_{IE} and a pair of NS_{AE}- NN_{AE} who were not part of the final group.

Each speaker in a pair received the appropriate version of the map based on his or her assigned role: instruction provider or instruction receiver. Providers were given the map with the route included and were tasked with verbally describing the route drawn on his or her map to the receiver. The receiver, in turn, attempted to reproduce this route on his or her map. These map tasks are intended to create spontaneous conversational exchanges and facilitate adaptations and modifications in each speaker's speech in response to the interlocutor. To further encourage verbal interaction, slight variations were present in each speaker's map version such as a duplicate or missing landmark item. To avoid any confusion, participants were notified that their maps may have slight variations.

3.4 DESIGN AND PROCEDURE

Each pair of speakers was assigned to one of the following groups based on their language background:

1. Native language group (NS_{AE}- NS_{AE})
2. Mixed dialect group (NS_{AE}- NS_{IE})
3. Mixed language group (NS_{AE}- NN_{SP})

To avoid social dominance phenomena associated with mixed-sex pairs, this study featured a gender-matched design (Namy et al., 2002). Each group had 24 speakers, 6 male-male pairs and 6 female-female pairs. Participants were assigned the role of

provider or receiver prior to arrival at the lab. These roles were retained through the course of the experiment and were counterbalanced for speaker pairs. For example, in the mixed dialect group of female pairs, three NS_{AE} were assigned the role of provider and the other three were assigned the role of receiver. The same was the case for both sexes in all language conditions. Members of the first pair in each group were assigned roles randomly via a coin flip. Subsequent pairs within the group were assigned roles that were opposite of the roles of the previous pair. Once assigned, receivers and providers did not switch roles across maps in the experiment.

Consent was obtained before any recordings or demographic information were collected. To ensure no bias during the experiment, convergence was not mentioned until the debriefing. Instead, participants were informed that the study examined the effect of practice on cognitive tasks (*i.e.* completing the map task).

Once participants had completed the language background questionnaire, each participant was recorded individually reading the vowel list and recording paragraph as the pre-task items. Then, they were seated in the booth with the provider backing the receiver (the receiver was seated at a table with the provider on his or her right hand side) and recorded completing the map task for the during-task data. Participants did not have access to each other's maps except through verbal descriptions. In order to reproduce a provider's map route onto a receiver's, providers offered verbal instructions on how to reproduce the trail drawn on their map. Participants were encouraged to talk as naturally as possible and gesture as necessary but they were asked not to draw paths in the air. As noted above, participants were notified that minor variations such as duplications and omissions exist in the provider and receiver versions of the maps. They were instructed not to talk simultaneously to avoid overlapping speech. Receivers were requested to summarize the path once each map was completed. This was done to elicit more balanced

speech from both participants. After the map task was complete, each participant recorded the vowel list and recording paragraph one last time to create post-task data. Once all recordings were complete, the participants were told that the actual goal of this study was to examine phonetic convergence in vowels and rhythm. Lastly, participants signed a debriefing form before leaving. All subjects were paid $10/hr for their time.

Even though participants were not told there was a time limit on the task, dyads that did not complete all four map routes within an hour were stopped. This was done not only to limit the amount of time each dyad took in the lab for fear of fatigue but also to ensure that all recordings were completed within two hours for each dyad due to scheduling requirements.

3.5 ANALYSIS AND MEASURES

This study is an analysis of speech production changes in dyads as they participate in a map task; the native language group provides data for phonetic convergence between speakers of the same variety of a language (English), the mixed dialect group provides data for PC between speakers of different dialects of a language and the last group (mixed language) of native and non-native speakers of English provides data on native-non-native interactions. For each group, speech tokens (utterances and vowels) extracted from scripted recordings and spontaneous speech were used for acoustical and statistical analysis. Data provided by vowel lists were used in the segmental analysis detailed below. Sentences from the recording paragraph and map task were used in the suprasegmental analysis of rhythm. Following Babel (2009a), male and female speakers within each language/dialect group were analyzed separately in all analyses. This was done for two reasons. Since the main goal of the study was to examine

the effect of language background on PC in vowels and rhythm, the effect of sex on PC was not directly examined. This allowed for the size of the statistical model to be reduced and more easily interpreted (see below for all variables examined). Furthermore, keeping the female and male analyses separate avoids obscuring any effects of convergence due to sex differences in acoustical properties. This dissertation examines each of the language conditions as individual experiments (described in Chapters 4, 5 and 6). In the following chapters, the terms 'convergence' or 'divergence' will be used to specifically refer to the type of PC that is noted. To generally refer to a change in a dyads speech, 'adaptation' will be used interchangeably with PC.

The mixed dialect group is useful in demonstrating how large a formant model with both sexes would be. The dependent variables (DV) are F1 or F2 while the independent variables (IV) are vowel (æ, ɑ, i, u, e, o) and task (pre or post), role (provider or receiver), dialect (IE or AE), creating a 6 X 2 X 2 X 2 ANOVA. Including gender would have increased the number of levels and possible interactions by 15. Language conditions were analyzed separately to simplify the model as well as to keep subject coding consistent. For example, including all males in a single analysis would have created a 6 X 2 X 2 X 3 X 3 model with F1 or F2 as DV and vowel (æ, ɑ, i, u, e, o) and task (pre or post), role (provider or receiver), language spoken (IE or AE or SP) and language condition (NS_{AE}-NS_{AE}, NS_{AE}-NS_{IE}, NS_{AE}-NN_{SP}) as IV. Furthermore, speaker pairs in the NS_{AE}-NS_{AE} condition are numbered 1-6. However, speaker pairs in the NS_{AE}-NS_{IE} and NS_{AE}-NN_{SP} conditions are numbered 1-3 because three pairs are set up in the $Provider_{AE}$-$Receiver_{IE/SP}$ and three are set up in the $Provider_{IE/SP}$-$Receiver_{AE}$ configuration. A model this large would make any intelligent interpretation of higher order interactions virtually impossible. Dividing the model in this manner makes it less unwieldy and more

amenable to interpretation. This restricts the interpretation in a manner that does not allow sex effects or language background effects to be compared directly.

'Speaker' was considered the within-subjects error term in the vowel analyses. But because EMS is adept at detecting idiosyncratic rhythm, it was expected that within each pair, idiosyncratic rhythm would contribute to adaptations and so 'speaker pair' was included as a factor in rhythm analyses. A nested factorial ANOVA was used to detect rhythm convergence with sentence token used as the error term. Thus, effects that were specific to a dyad were noted in rhythm analyses but not in vowel analyses.

Speakers were asked not to talk at the same time but in the event that overlapping speech did occur, it was excluded from the analysis. Segmentation was carried out using the rules outlined in Peterson and Lehiste (1960) using Praat (Boersma & Weenik, 2003), which was also used for all acoustical measurements and analysis. Statistical analysis was carried out using Matlab (2012b) and R (R Core Team 2012). Certain plots were created using the ggplot2 package and hierarchical linear models used in Chapter 7 were created using the nlme package in R. Because male and female data were analyzed separately, significance was alpha adjusted using Šidák correction to 0.025. Subsequent analysis of simple effects were then alpha adjusted based on the relevant factor. For example, to analyze a subsequent effect of vowels, significance would be evaluated at $\alpha = 0.004$.

3.5.1 Segmental analysis

Vowel edges were marked at the first and last glottal pulse of each vowel by hand. Using these demarcations, the 20%, 50% and 80% points were calculated via a Praat script written specifically for this study. For any values that fell outside the range of the expected values based on Hillenbrand et al., (1995) and Peterson & Barney (1952), hand

checking and correction were performed. F1 and F2 values were subjected to separate Analysis of Variance (ANOVAs). MANOVAs were not attempted because the correlation between F1 and F2 was not significant. Since it is expected that F1 and F2 will detect vowel differences, main effects of vowels from the formant analysis are reported but not discussed. Rather than marking vowels based on the perceived vowel or F1/F2 values, vowels were marked as the intended target vowel. Thus, if a speaker pronounced 'hod' as 'hawd' instead of 'hɑd', it was still marked as 'hɑd' to make coding consistent. This difference was a dialectal one and noted for IE speakers.

Because formant measurements are the standard metric to distinguish vowel quality, they were used in this study to measure PC in vowels. VISC, Euclidean distance and interlocutor similarity (IS) were also used to measure vowel PC. Though Euclidean distance has been used before (Babel 2009a; Pardo et al., 2010), VISC and IS are novel to this study of PC. Vowel similarity was fairly high due to overlap in vowel inventories across dialects and languages. This led to a uniform distribution of the metric creating a violation of the assumption of a normal distribution for ANOVAs. Thus, the f-value of the ANOVA for the IS measures was subjected to a bootstrap analysis using sampling with replacement. Because of the way the IS measures are calculated (see below), similarity between a provider-receiver pair is quantified by a number. Thus, role and linguistic background cannot be considered as factors in these analyses. Because female and male dyads were analyzed separately, instead of 95% confidence intervals (CI), 97.5% CI were used to avoid type I error.

3.5.1.1 Midpoint F1 and F2 and VISC

F1 and F2 were noted for all vowels that are the focus of this study. Three measurements were made for each F1: Onset, offset and center frequencies were noted at 20%, 50% and 80% of the vowel duration respectively. The center or midpoint, frequencies of F1 and F2 were used as the dependent variables (DV) in an analysis with role, linguistic background and task as independent variables (IV). The difference between the onset and offset was used to calculate the onset + offset measure of VISC. A separate analysis was used to gauge dynamic vowel changes taking place during convergence using VISC as the DV and the same IVs as the static vowel analysis.

3.5.1.2 Interlocutor similarity (IS)

IS is a novel approach to measuring vowel PC. Cosine similarity, which is the basis for this measure, is frequently used to measure varied topics such as semantic similarity in documents and facial verification (Nguyen, 2011). IS was calculated for each speaker pair's vowels using this cosine similarity metric. Cosine similarity is measured as the cosine of the angle between two vectors and calculated using the formula: $A.B / \|A\| \|B\|$. Figure 3.1 shows the difference between cosine similarity and Euclidean distance. It can be seen from the diagram that if Euclidean distance is the distance between two points, cosine similarity is the angular relationship between them. If two vectors are maximally different or orthogonal, the cosine similarity value for them is 0. If, on the other hand, the two vectors are identical, the cosine similarity value for them is 1. Like Euclidean distance, this measure is capable of simultaneously tracking similarity increases or decreases in F1 and F2. Cosine similarity is more sensitive than the Euclidean distance because it tracks orientation as well as distance in F1-F2 space ($\cos(0) = 1$ but $\cos(180) = -1$). For example, cosine similarity would detect if a receiver's

vowels were raised from pre-task to post-task even if the Euclidean distance between the receiver's and the provider's vowels remained unchanged. For the current experiment, each vowel was assigned a vector consisting of its midpoint F1 and F2 values. Cosine similarity between each provider-receiver pair's vowel vectors was measured in two ways: on the entire vowel space (*i.e.* across the 6 vowels) as well as on a per vowel basis.

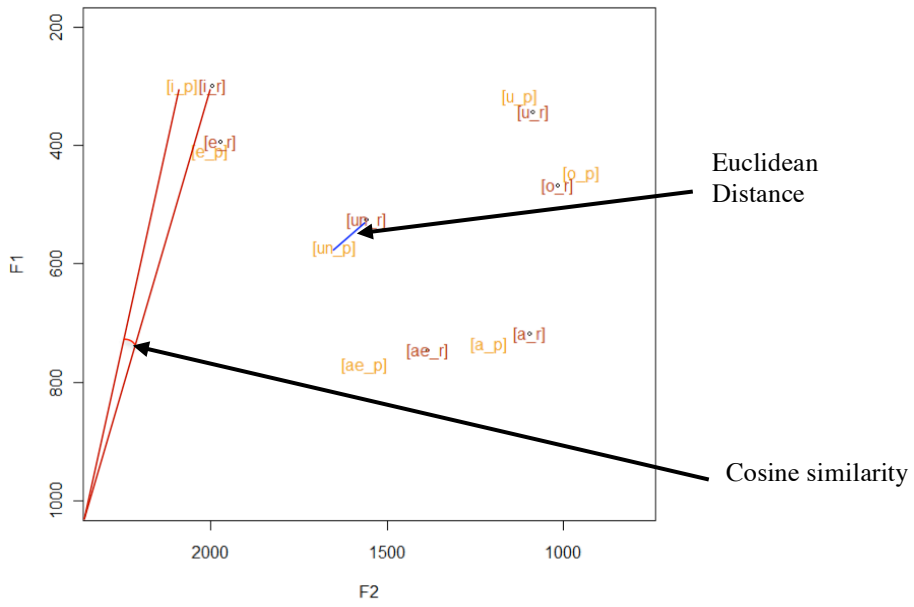


Figure 3.1: Diagram showing the difference between Euclidean distance and cosine similarity. Cosine similarity is the angular distance between two or more points. i_p and i_r denote the provider's and receiver's i vowel.

The first measure is the systemic IS or sys-IS. Sys-IS is similarity between a provider's and a receiver's vowel set (*e.g.* the similarity between a provider's and a receiver's entire vowel space, /æ, ɑ, i, u, e, o/). While sys-IS was more sensitive to systemic changes in vowels; the second measure, specific IS or spec-IS, was expected to be more sensitive to vowel specific convergence. Spec-IS was obtained by calculating the

cosine similarity between a provider's and a receiver's vowels on a per vowel basis (*e.g.* the similarity between a provider's and a receiver's /o/). Thus, each pair received two sys-IS scores (pre-task and post-task) as well as 12 spec-IS scores for 6 pre-task and 6 post-task vowels. These scores provide a measure of vowel-general and vowel-specific midpoint similarity for each speaker pair. Systemic and specific IS values were used as DVs in separate statistical analyses using ANOVAs. Both measures were included because it was uncertain which approach would be more informative.

3.5.2 Suprasegmental analysis

The EMS + centroid approach was used to extract utterance-level spectral rhythm information from the sentences in the recording paragraph and maps. Using Liss et al.'s (2010) procedure, EMS was used to extract the power spectrum of each landmark item phrase. A flowchart that outlines the steps involved in extracting the EMS is provided below in Figure 3.2.

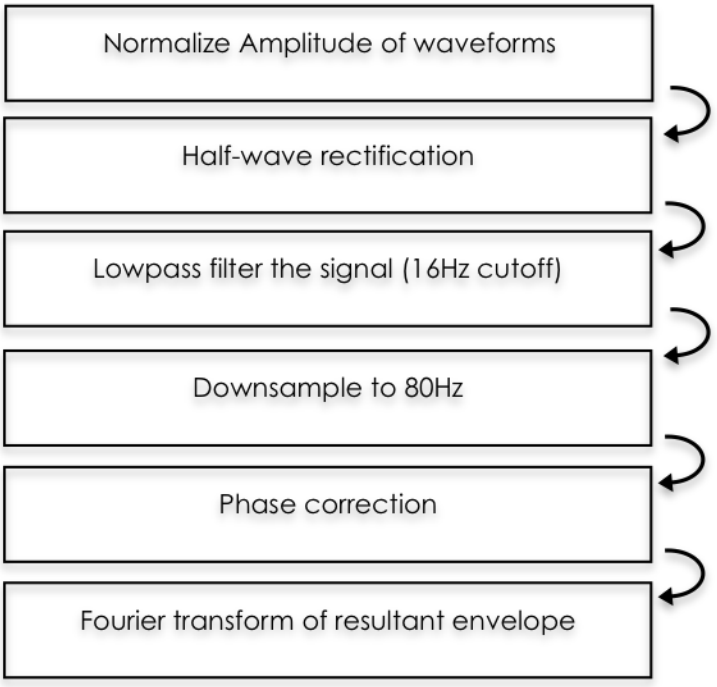


Figure 3.2: Steps involved in extracting the EMS of a sound wave.

Figure 3.3 shows the waveform, spectrogram and EMS of the word 'bumblebees' as spoken by a male speaker of AE. This is the power spectrum (amplitude vs. frequency) of speech filtered with a lowpass cutoff of 16Hz. It can be seen from the bottom diagram that EMS represents low-level frequency modulations of the speech. This measure captures rhythmic variability as marked by amplitude undulations due to vowel and consonant distributions. Liss et al. (2010) and LeGendre et al. (2009) calculated EMS for the entire signal and for each octave band through 8000 Hz. Filtered speech was used to mimic the filtering of a cochlear implant. For the purposes of this study, which does not involve any clinical subjects, EMS was calculated over the entire signal.

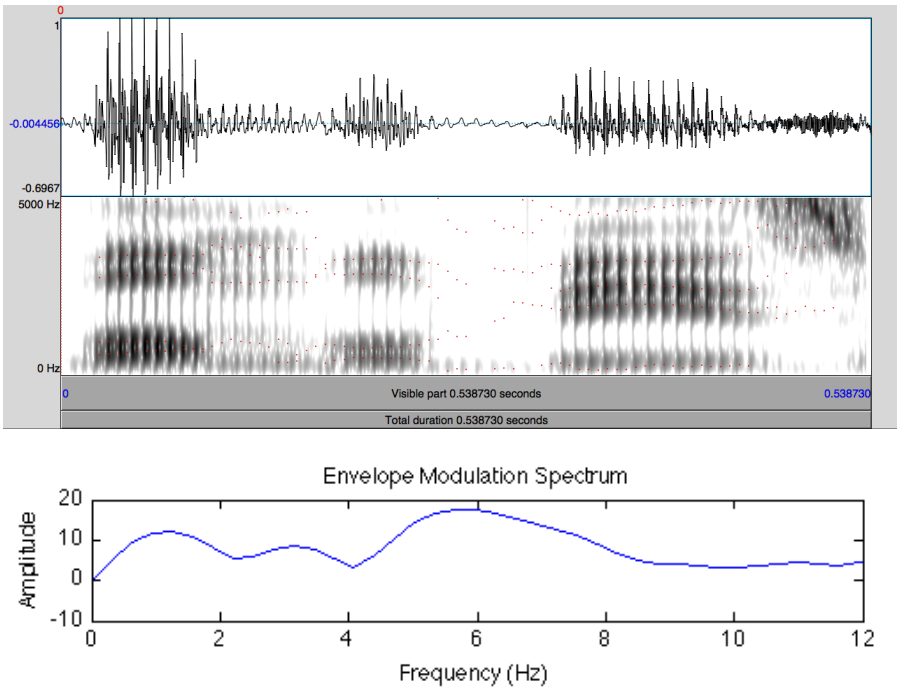


Figure 3.3: Waveform, spectrogram and EMS of *'bumblebees'* (spoken by male AE speaker)

Modulation frequencies in the range 0-16 Hz contribute to sentence intelligibility (Drullman, Feston & Plomb, 1994). Keeping this in mind, the cutoff frequency for low band pass filtering used in this study was 16 Hz instead of 30Hz, used by Liss et al. (2010) and LeGendre et al. (2009), and 10 Hz, used by Tilsen and Johnson (2008). This created a signal that retained the acoustic characteristics of its components (such as onsets and offsets) but made it incomprehensible. Previous research has used peak tracking to measure rhythm (Liss et al., 2010; LeGendre et al., 2009; Tilsen & Johnson, 2008). EMS tracks periodic components in speech. Thus, a signal repeating at a given period will demonstrate a peak at 1/period Hz. Since speech is more complicated, it will have several relevant peaks in the EMS domain. As speakers in the current study were from one of three rhythmically distinct backgrounds, it is likely that more than a single

peak would contribute to each speaker's rhythm. This was supported by pilot research that found instances of EMS with several near equal peaks. To ensure that spectro-temporal information provided by less prominent peaks is not lost, this study uses EMS + centroid to calculate rhythm (Rao & Smiljanic, 2011). The centroid measure has been used in vastly different fields such as phonetics to distinguish between the fricatives, /s/ as in '**s**it' and /ʃ/ as in '**sh**oe' (Beckman, Yoneyama & Edwards, 2003) and in music to distinguish between timbre of tones (Kendall, 2002). The spectral centroid is a weighted mean of the spectrum and is calculated using the following formula (step = 1 Hz):

$$Centroid = \frac{\sum_{n=0}^{N-1} Frequency(n)Amplitude(n)}{\sum_{n=0}^{N-1} Amplitude(n)}$$

Nested factorial ANOVAs allow for each sentence token to be used as a data point and were used to detect rhythm PC. In this analysis, speakers were expected to have differing idiosyncratic rhythm and were treated as a fixed factor. In this case, the sentence token (each sentence from the recording paragraph) was the random factor. Thus, the centroid of the spectral rhythm was used as the DV in analysis using nested factorial ANOVAs.

Rhythm convergence could be realized in one of two ways: a decrease in the difference between the centroids of a dyad's provider and receiver or a decrease in the magnitude of the centroid. A higher EMS centroid denotes increased variability in speech due to speaking style, syllable distribution, pauses or dysfluencies (Rao & Smiljanic, 2011). A reduction in the difference between a provider-receiver pair's centroids would indicate listener-specific adjustments at the rhythm-level. In addition to a reduction in the difference between a dyad's centroids, convergence may also be denoted by a decrease in

the height of the centroid denoting a reduction in rhythmic variability. A lower post-task centroid would suggest a reduction in within-speaker speech rhythm variability as two speakers converge on a mutually similar, and less variable, rhythm pattern. Conversely, divergence would be represented by an increase in the distance between a provider and receiver's centroids or an increase in the height of both their centroids.

I examine rhythm during the map tasks and the role of explicit imitation in convergence in separate chapters. There is a lack of perception studies that examine human responses to the correlates detected by EMS. To ensure controls on the mode of presentation and recording (citation from the recording paragraph vs. spontaneous speech from the map task), map task data were included in a separate analysis (Chapter 7).

Chapter 4: Native Language Group (NS_{AE}-NS_{AE}, Close Condition)

4.1 INTRODUCTION

The purpose of the current experiment was to test vowel and rhythm specific PC in speaker pairs that shared the same native language, specifically AE. Although vowels have been studied with respect to PC, rhythm has been largely unexplored in this domain. This experiment adds to the current knowledge pool by examining PC in rhythm using a spectral measure of rhythm rather than the previously used duration measurements. Additionally, IS measures were used in this experiment in an attempt to quantify vowel-systemic and vowel-specific changes in a speaker pair's vowels. Dialects within AE (regional or sub-dialects) were not restricted to any particular variety though speakers were most likely to speak with a general AE accent with some southern features. This was done in particular to maintain consistency with the other two language conditions, where it was not feasible to control for dialectal variation. More information about NS_{IE} and NN_{SP} subjects is provided in Chapters 5 and 6.

In interactions involving speakers that share the same native language, PC has been demonstrated in various segmental features such as vowel formants, spectra and VOT (Pardo, Jay, & Krauss, 2010; Nielsen, 2008; Delvaux & Soquet, 2007). For example, Delvaux and Soquet (2007) showed that female speakers of the Mons and Liège dialects of Belgian French converged in both consonants and vowels with the other dialectally distinct speech for sounds that are produced differently in these two dialects: /p, o, i, s/. In another study, Babel (2009a) used a word-shadowing task to show that both men and women converge to the vowels of 'model speakers'. Nielsen (2008) examined VOT in stops produced by AE speakers during a word-shadowing task. Not only did speakers imitate artificially lengthened VOTs in stops, but they also generalized the

convergence from the /p/ token, on which they were trained, to a new phoneme, /k/, suggesting that both word-level and feature-level properties are susceptible to PC.

Perceptual tests of interactions between speakers of the same L1 also reveal convergence. Kim et al. (2011) found that speakers who share an L1 and regional dialect demonstrated convergence in their speech. They also found that speakers who share an L1 but not the same regional dialect did not converge. However, multiple studies that do not control for regional reported task-induced adaptations (Pardo, 2006; Babel 2009a; 2010; Pardo et al., 2010; Nielsen, 2008).

Only recently has research focused on the effect of convergence on rhythm (Krivokapic, 2013; Rao et al., 2011c). Krivokapic (2013) examined convergence of rhythm in four sex-matched dyads of IE and AE because they are considered rhythmically distinct. AE rhythm is considered stress-timed, whereas IE is described as syllable-timed. Krivokapic (2013) found that one out of four IE speakers' rhythm altered to the more AE style stress-timed pattern after the interaction. Rao et al. (2011) suggested that linguistic background affects convergence in dyads comprising of native and non-native speakers of AE. They found that male and female speakers who spoke the same national variety of English (specifically, AE) either natively or non-natively were more likely to converge in rhythm than those that spoke different varieties of the same language (AE and IE). Their results suggested that speakers who used the same linguistic rhythm natively or non-natively would converge and those that spoke with different linguistic rhythm would diverge.

4.2 HYPOTHESES

4.2.1 Vowels

Overall, Convergence in vowels would be detected via receiver's and provider's vowel formants, VISC and IS measures becoming more similar post-task compared to pre-task. In formants and VISC, convergence would be noted by a reduction in the difference in provider and receiver post-task values. Sys-IS and spec-IS would indicate convergence via an increase in their post-task values. Generally, sys-IS would detect convergence across dyads' vowel systems. If vowel-specific convergence was detected, it would be via F1 and F2 and spec-IS but not sys-IS which has been specifically designed to ignore vowel specific changes.

4.2.2 Rhythm

The main prediction for rhythm was that convergence would be detected in dyads' speech patterns. Convergence would be characterized by either a decrease in the distance between the speaker pair's centroids or a reduction in the height of both the provider and the receiver's centroids. Considerations for evaluating convergence and divergence using EMS + centroid are outlined in Section 3.5.2.

EMS is particularly suited to detect idiosyncratic and sex differences in rhythm (*e.g.* LeGendre et al., 2009). Thus it was also predicted that speakers would demonstrate individual differences in rhythm PC.

4.3 METHODOLOGY

Methods and stimuli are as described in the chapter on methodology, Chapter 3. Information specific to the native language group is provided below.

4.3.1 Participants

The native language group included 24 speakers (12 male). Their ages ranged from 17-44 years old (mean = 25.14 years, SD = 7.92). Participants did not have any known speech or hearing impairments at the time of recording and were native speakers of AE. Participants were either students (graduate or undergraduate) at the University of Texas at Austin or professionals living in the greater Austin area. Six of the speakers had lived in a state other than Texas, including Massachusetts, Colorado, California, Maryland, Illinois and Ohio, for at least a year. All but six had some experience with a second language (Spanish, French, Mandarin, Japanese or German) via high school or college courses as part of a language requirement but were not fluent in any of these languages as indicated in the background language questionnaire. Participants took an average of 25.50 minutes (SD = 0.46) to complete all four maps in the map task. Due to time limitations, one male pair was stopped after 50 minutes after completing three of the four maps.

4.3.2 Rhythm

The first three sentences of the paragraph were discarded due to clipping in one speaker's recordings. Thus, a total of 8 sentences from each speaker for each task (pre and post) were included in the final analysis for rhythm.

4.4 RESULTS

The following section describes the results of the formant midpoint, IS and rhythm analyses in that order. For each section, the descriptive trends are discussed first followed by the statistical findings separated by sex. Alternative descriptive plots for vowels are also provided in Appendix E. Results from all statistical analyses are provided in Appendix D.

4.4.1 Midpoint formant analyses

4.4.1.1 Female speakers

Figure 4.1 shows the average formant values for all six vowels separated by role and task for female speakers. These plots do not reveal any vowel-specific or general trend in terms of task. It seems that speakers tended to maintain the difference in F1 and F2 with respect to their partner after the map task. For example, in /ɑ/, both F1 and F2 were larger[1] for receivers than for providers.

[1] 'larger' and 'smaller' are used to indicate F1 and F2 values to disambiguate them from 'high' and 'low' which can also mean vowel quality.

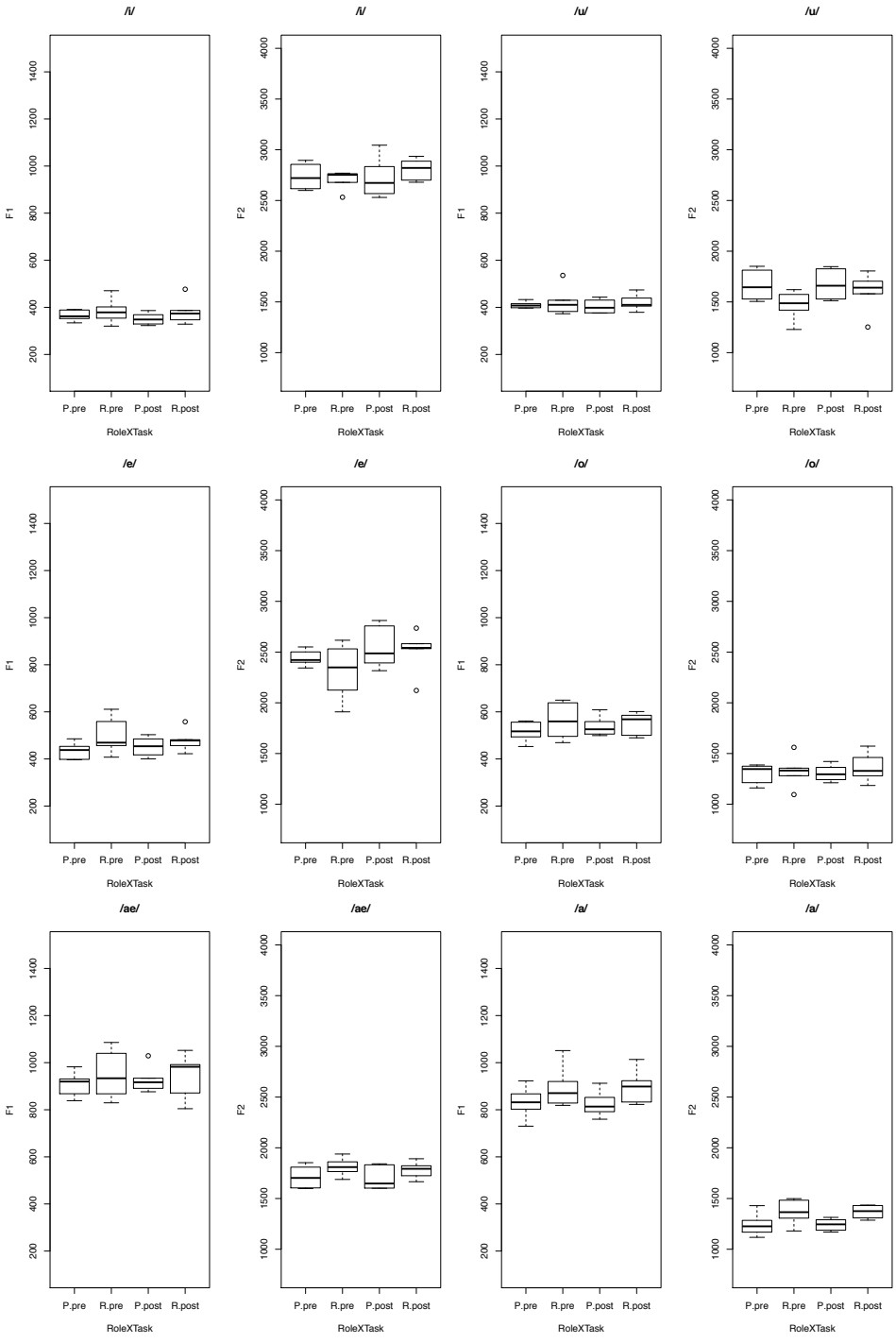


Figure 4.1: Average F1 and F2 values for all female vowels separated by role and task. Vowels are arranged in order of vowel height from top to bottom (/ɑ/ and /æ/ are listed as ‘a’ and ‘ae’ respectively in figure).

Mean female F1 and F2 values based on role and task are provided in Table 4.1. All standard errors (SE) are listed in parentheses. These data show that female receivers and providers increased their F2 values from pre-task to post-task. Providers increased their F1 values from pre-task to post-task.

	Providers		**Receivers**	
	Pre-task	**Post-task**	**Pre-task**	**Post-task**
F1	576.509 (36.56)	582.977 (37.19)	614.951 (39.33)	612.194 (39.24)
F2	1838.760 (95.48)	1862.570 (99.13)	1820.328 (90.71)	1899.950 (96.84)

Table 4.1: F1 and F2 (Hz) means for female speakers separated by role and task (SE in parentheses).

A 6 X 2 X 2 mixed-design ANOVA, using F1 as a dependent variable (DV) and vowel type (/æ, ɑ, i, u, e, o/) and task (pre or post) as within subjects factors and role (receiver or provider) as a between subjects factor was run on the female speakers' vowels. At $\alpha = 0.025$ (alpha adjusted), it revealed main effects of vowel ($F(5, 25) = 527.87$, $p < 0.01$) and role ($F(1, 60) = 14.78$, $p < 0.01$). No other main effects or interactions were noted. Providers had smaller F1 values indicating more raised vowels than receivers (Figure 4.2).

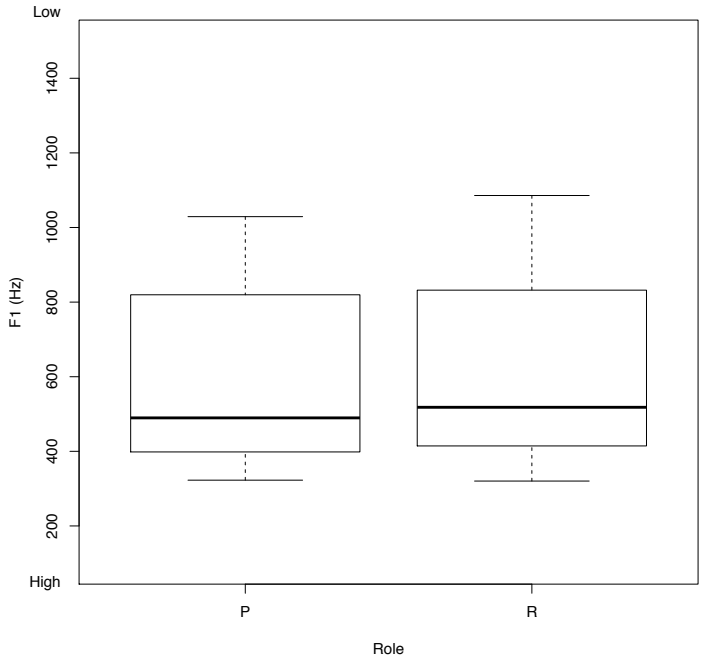


Figure 4.2: Significant F1 values separated by role (female speakers). 'High', and 'low' indicate vowel quality.

A separate analysis using the same IVs but F2 as the DV revealed a main effect of vowel ($F(5, 25) = 378.13$, $p < 0.01$) and an interaction between vowel and role ($F(5, 60) = 3.52$, $p < 0.01$). Post hoc tests of simple effects at significance level 0.004 determined that providers produced smaller F2 values for /ɑ/ than receivers (Figure 4.3).

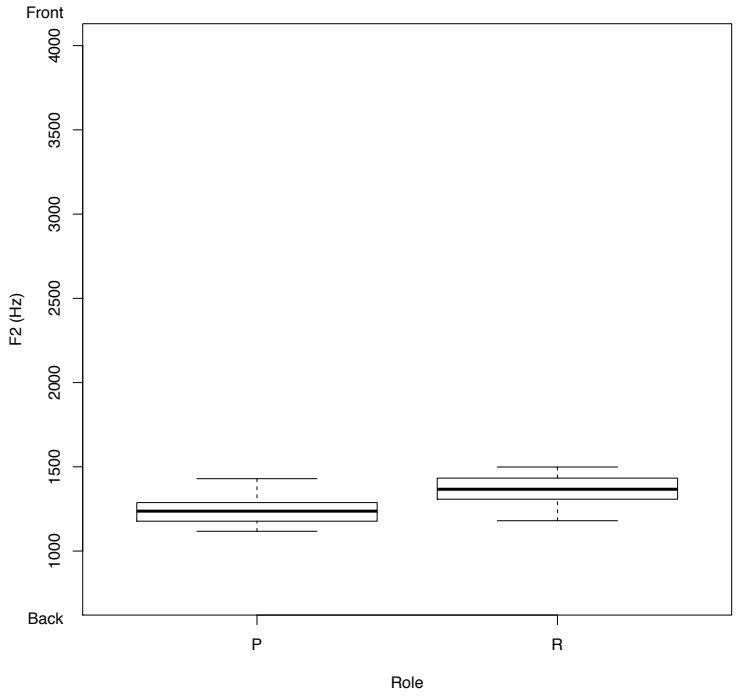


Figure 4.3: Significant F2 values for /ɑ/ separated by role (female speakers). 'Front' and 'back' indicate vowel quality.

4.4.1.2 Male speakers

Figure 4.4 shows average formant values for all six vowels separated by role and task for male speakers. These plots do not reveal any vowel-specific or general trend in terms of task. Receivers appear to have larger F1 and F2 values than providers. For example, male speakers' F2 value for /ɑ/ may have been larger for receivers than for providers.

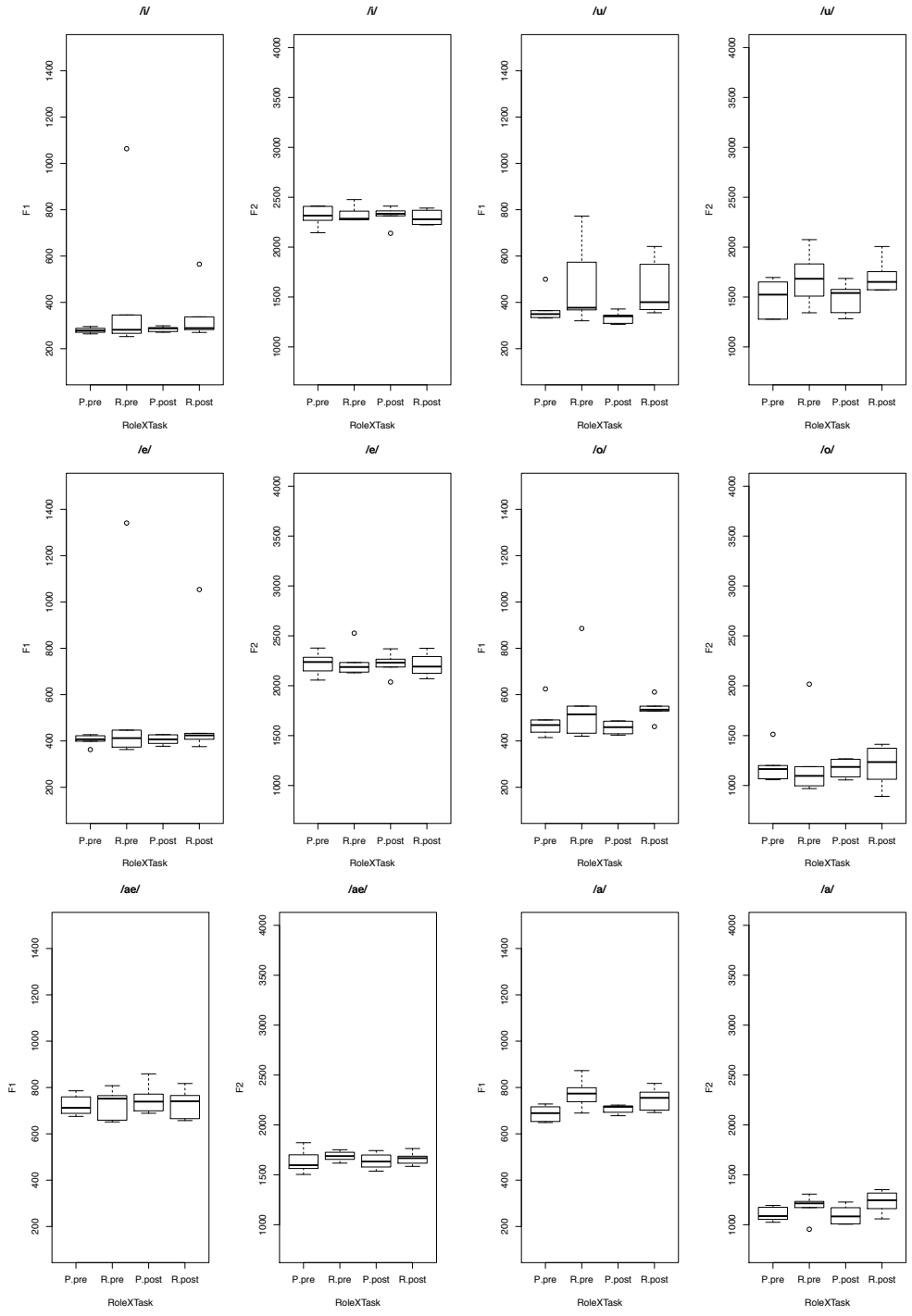


Figure 4.4: Average F1 and F2 values for all male vowels separated by role and task. Vowels are arranged in order of vowel height from top to bottom (/ɑ/ and /æ/ are listed as ‘a’ and ‘ae’ respectively in figure).

Mean male F1 and F2 values based on role and task are provided in Table 4.2. All standard errors (SE) are listed in parentheses. These data show that male speakers who were receivers lowered their F1 values and F2 values from pre-task to post-task. Providers do not show any notable task specific changes. Compared to the providers, receivers also showed larger F1 and F2 values.

	Providers		**Receivers**	
	Pre-task	**Post-task**	**Pre-task**	**Post-task**
F1	489.703 (28.44)	491.982 (30.41)	589.242 (41.79)	555.888 (31.93)
F2	1668.123 (81.09)	1670.829 (81.10)	1724.873 (81.26)	1715.025 (74.49)

Table 4.2: F1 and F2 (Hz) means for male speakers separated by role and task (SE in parentheses).

Two separate mixed-design ANOVAs using the same IVs as reported for the female analysis above and F1 and F2 as DVs were run on the male data. The F1 ANOVA revealed main effects of role ($F(1,60) = 12.57$, $p < 0.01$) and vowel ($F(5,25) = 32.08$, $p < 0.01$). No interactions were noted. An analogous ANOVA with F2 are the DV and the same IVs as above revealed main effects of role ($F(1,60) = 7.32$, $p < 0.01$) and vowel ($F(5,25) = 180.19$, $p < 0.01$). Mean provider and receiver F2 values are provided in Table 4.2, which show that providers had smaller F1 and F2 values than receivers. Mean provider and receiver F1and F2 values are plotted in Figure 4.5.

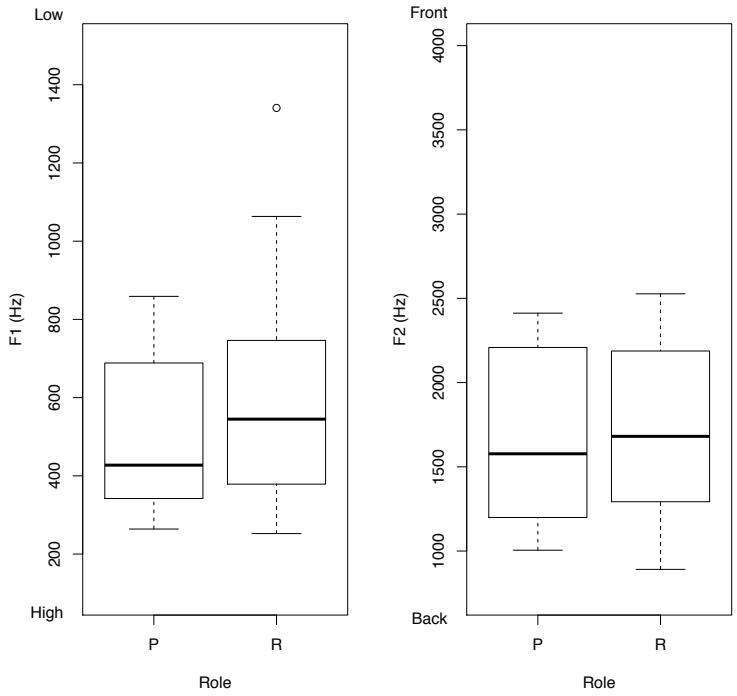


Figure 4.5: Significant F1 and F2 values separated by role for male speakers. 'Low', 'high', 'front' and 'back' indicate vowel quality.

4.4.1 Interlocutor similarity (IS)

The mean values for both IS measures between female providers' and receivers' vowels are provided in Table 4.3. The increase in similarity scores for both systemic and specific IS demonstrates a decrease in distance between the vowel midpoints of each dyad's receiver and provider. These values suggest that sys-IS increased after the interaction for all speakers. Spec-IS showed an increase for female speakers from pre-task to post-task. The men maintained their spec-IS values from pre-task to post-task.

	Pre-task	Post-task
Female sys-IS	0.993 (0.00)	0.996 (0.00)
Female spec-IS	0. 998 (0.00)	0. 999 (0.00)
Male sys-IS	0.984 (0.00)	0.988 (0.00)
Male spec-IS	0.996 (0.00)	0.996 (0.00)

Table 4.3: Means of pre-task and post-task systemic and specific ISs (SE in parentheses are 0 up to two significant digits).

Bootstrap analyses of two repeated-measures ANOVA with sys-IS as DV and task (pre, post-task) as IV were run separately on the male and female data. Recall that sys-IS and spec-IS values formed a uniform distribution. As mentioned in Chapter 3, sys-IS and spec-IS are a measure of similarity between the receiver and provider. As a result, role cannot be tracked anymore. These bootstraps revealed significant effect of task on sys-IS for the female speakers. A small effect of task was noted for male speakers. The F-value CI is so close to 1 that it indicates this effect was just barely significant. In order to avoid a spurious result, this interaction will not be discussed as meaningful. The 97.5% (equivalent of $\alpha = 0.025$) CI for the sys-IS task F-values for female and male pairs are provided below in Figure 4.6.

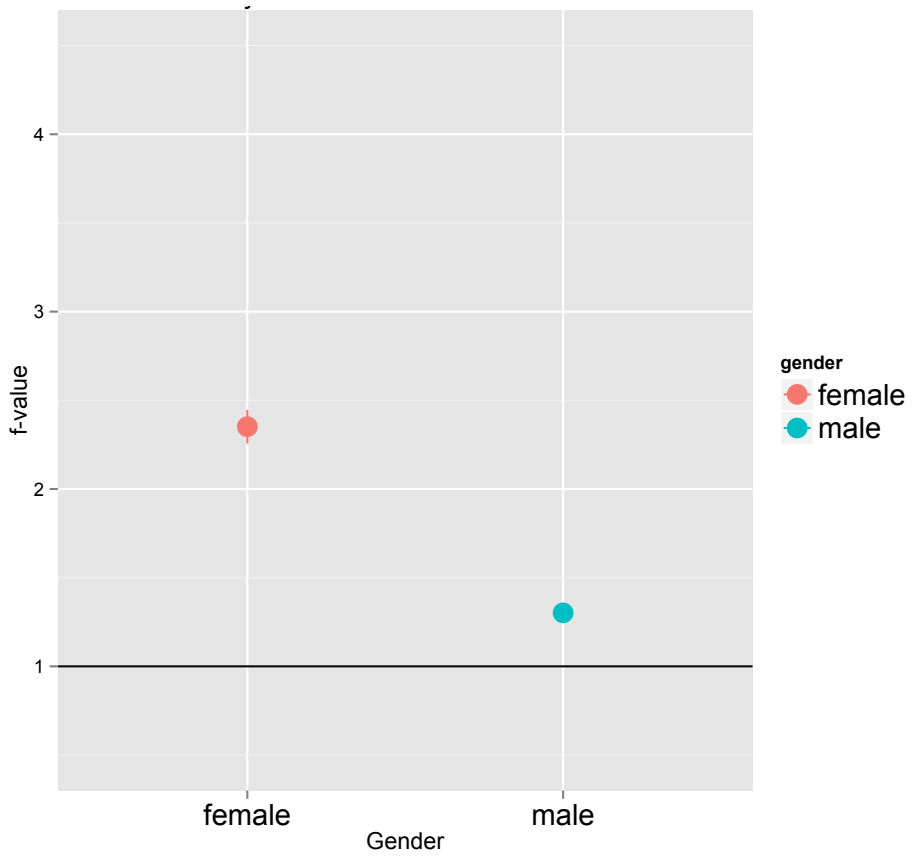


Figure 4.6: 97.5% confidence interval of the F-values for the main effect of task on systemic IS for male and female vowels. Red points indicate female and blue indicate male f-values.

In addition to the sys-IS, two other bootstrap analyses with spec-IS as DV and task and vowels as IVs were also run. These showed a main effect of task for the female speakers but no interaction between task and vowel. There appears to be an interaction between vowel and task for the male data. However, this CI is close to one and for the reasons stated for sys-IS above, it won't be discussed as meaningful. The 97.5% CIs for the spec-IS task F-values for male and female speakers are provided below in Figure 4.7.

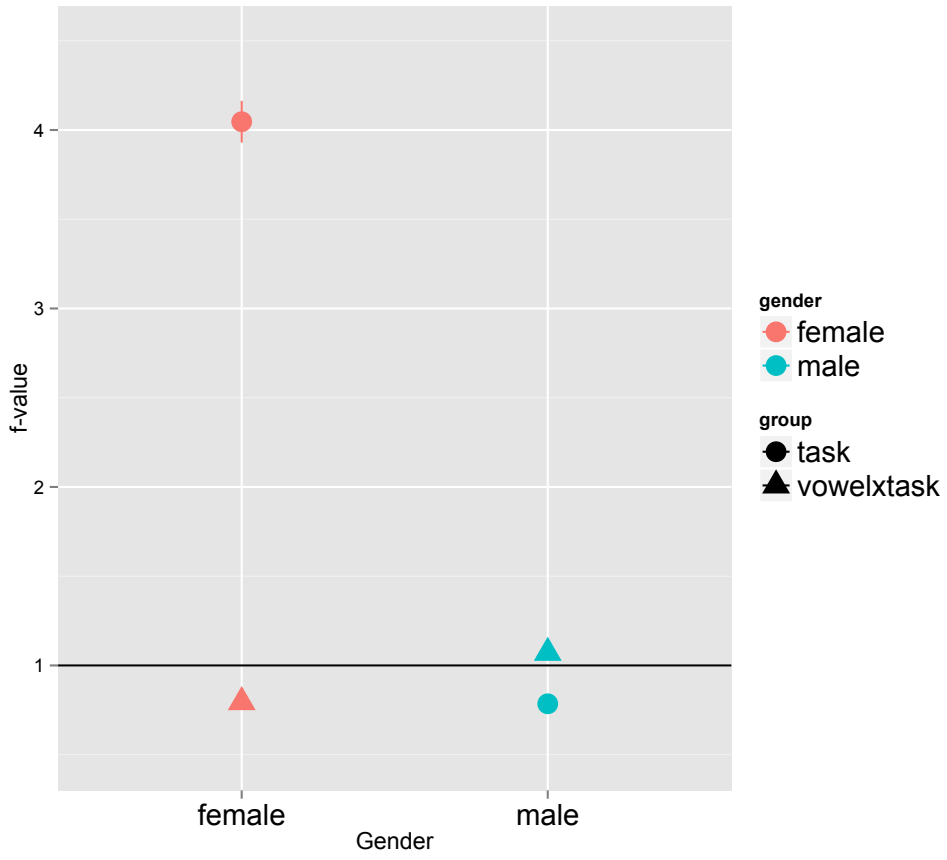


Figure 4.7: 97.5% confidence interval of the F-value for the main effect of task on spec-IS for male and female vowels. Red points indicate female and blue indicate male f-values. Circles indicate task f-values whereas triangles indicate f-values for the vowel X task interaction.

4.4.3 VISC

The analysis using VISC yielded no significance beyond an effect of vowels and was dropped from further analysis and in all subsequent experiments. The reasons why VISC did not detect any PC are discussed further in the general discussion in Chapter 9.

4.4.4 Rhythm

4.4.4.1 Female speakers

Female centroid averages are provided in Table 4.4. These averages suggest that women lowered their centroids slightly from pre-task to post-task. Women reduced the distance between their centroids from 0.254 to 0.199.

	Providers	
	Pre-task	**Post-task**
Providers	7.943 (0.28)	8.024 (0.23)
Receivers	8.197 (0.26)	8.223 (0.26)

Table 4.4: Means of rhythm centroids for female speakers separated by role and task (SE in parentheses)

The female data were subjected to a 2 X 6 X 2 nested factorial ANOVA that used role (receiver or provider), speaker pair (speaker pairs 1-6) and task (pre or post) as IVs and EMS + centroid values as the DV. Recall that speaker pair is included as an IV for the rhythm analysis because idiosyncratic rhythm could influence the results; for the vowel analyses it is the error term. Significance was evaluated at $\alpha = 0.025$ (alpha adjusted for sex).

This analysis revealed a main effect of speaker pair ($F(5, 70) = 8.6703$, $p < 0.01$) and an interaction between role and speaker ($F(5, 70) = 4.0413$, $p < 0.01$). Post-hoc Bonferroni analyses at $\alpha = 0.004$ revealed that speaker pair 4 was significantly different from speaker pairs 1, 3, 5 and 6 (Table 4.5). Moreover, the receiver and provider of female speaker pair 5 had significantly different centroids. The provider had a lower centroid (mean = 7.754, SE = 0.44) than the receiver (mean = 9.141, SE = 0.45).

	Average centroid values
Speaker pair 1	8.110 (0.28)
Speaker pair 2	7.900 (0.26)
Speaker pair 3	8.429 (0.28)
Speaker pair 4	6.810 (0.29)
Speaker pair 5	8.447 (0.33)
Speaker pair 6	8.883 (0.32)

Table 4.5: Means of female rhythm centroids showing different rhythm centroids for all speakers (SE in parentheses).

4.4.4.2 Male speakers

All male centroid averages are provided in Table 4.6. These averages suggest that men lowered their centroid from pre-task to post-task. Male speakers reduced the distance between providers' and receivers' centroids from 0.434 to 0.093 due to the interaction.

	Providers	
	Pre-task	**Post-task**
Providers	9.253 (0.25)	8.643 (0.22)
Receivers	8.819 (0.20)	8.550 (0.23)

Table 4.6: Means of rhythm centroids for male speakers separated by role and task (SE in parentheses)

An analogous 2 X 6 X 2 nested factorial ANOVA was also used to examine the male data. It revealed a main effect of task ($F(1, 84) = 7.7480$, $p < 0.01$) and an interaction between speaker pair and task ($F(5, 84) = 2.7373$, $p < 0.05$). Post-hoc tests of simple effects using $\alpha = 0.004$ revealed pre-task centroids were not significantly different from post-task centroids for any specific speaker pairs at an adjusted alpha of 0.004.

Instead, speakers generally demonstrated lowered centroids after the task (mean = 8.59, SE = 0.16) than before it (mean = 9.03, SE = 0.16).

4.4 DISCUSSION

4.5.1 Vowels

F1 and F2 midpoint vowel analysis revealed significant differences due to role and vowel. Compared to female providers, female receivers fronted their /ɑ/ vowels. Male receivers fronted all six vowels when compared to male providers. Based just on the formant analyses, these effects are difficult to interpret. Role-based differences are quite pervasive in convergence (Pardo, 2006; Pardo et al. 2010; Kim et al., 2011). Kim et al. (2011) reported role variations that may be analogous to the current findings. They speculate that differences due to role may have been the result of speakers taking on 'leader' and 'follower' roles but do not discuss these findings any further. Because no effect of task was found in the formant analyses, the differences between providers and receivers are not due to convergence.

The IS scores, on the other hand, did reveal significant changes in vowel production due to the task. IS measures, which capture dyadic similarity (both vowel systemic and specific) noted increased similarity from pre to post-task in women's vowels. Spec-IS showed a main effect of task but no interaction between vowel and task suggesting that while there was a general increase in vowel similarity after the interaction, no vowel-specific adaptations were noted. Considered together, spec-IS and sys-IS scores both indicate that women exhibited an overall increase in vowel similarity suggestive of convergence. This suggests that the IS measures were more sensitive to the vowel modifications arising from the interaction than formant measures were. This

finding is discussed further within the context of the larger study in the general discussion.

4.5.2 Rhythm

Rhythm analyses revealed significant speaker-specific differences in women. Female speaker pairs demonstrated distinct rhythm patterns confirming previous findings that EMS is capable of detecting individual differences in rhythm (*e.g.* LeGendre et al., 2009). Role was also responsible for determining rhythmic characteristics in one female speaker pair. Female speaker pair 5's provider had a lower centroid than speaker pair 5's receiver. This difference between the provider's and receiver's rhythm was probably due to differences in speech patterns. The providers were most likely to use statements to provide instructions whereas the receivers asked more questions and made statements about the map routes. The differences in rhythmic patterns due to role may be a result of the intonation differences in the types of statements utilized by the speakers. Alternatively, it is also possible that the provider in this pair used clear speech during the entire task resulting in a less variable rhythmic pattern.

The results revealed a significant effect of task in men. Therefore, rhythm adaptation was noted in men whose rhythm centroids were lower post-task than pre-task. Rao & Smiljanic (2011) found that a reduced centroid (due to clear vs. conversational speech) denoted a decrease in variability in the speaker's rhythmic pattern. This was also noted by Tilsen & Johnson's (2008) study, in which the peak of the amplitude power spectrum was lower for citation style speech than for conversational speech. Both the lowered peak and the lowered centroid indicate a decrease in variability from conversational speech to a more clear speech style which is marked by less vowel and

consonant deletions, longer vowel durations and fewer fluctuations in amplitude. These results suggest that male dyads converged by reducing their rhythmic variability. One manner in which they might have achieved this is by switching to a more clear way of speaking during the interaction.

The difference in male centroids dropped from 0.43 to 0.09 and the difference in female centroids dropped from 0.26 to 0.20 (Table 4.4). This reduction in the differences in their rhythm suggests that along with male speakers, female speakers may also have been on the verge of convergence. Given that the analyses were conducted on eight sentences per speaker in pre- and post-task, it is possible that a larger set of sentences and a longer interaction would have led to greater adaptations in speech patterns.

4.5.3 General trends

Thus, both vowel and rhythm adaptations were noted in this set of speakers who shared the same national variety of a language specifically, AE. This is different from Kim et al. (2011) who found that only speakers who were closest in linguistic distance converged, but it is in line with other studies that showed adaptations in AE speakers whose regional dialects were not explicitly controlled (Babel, 2009a; Pardo 2006; Pardo et al., 2010). Women converged in vowels, as indicated by the IS measures but not in rhythm. Men on the other hand, showed convergence in rhythm but not in vowels. The reason for these differences is unclear. One possibility is that the small dataset is revealing differences that are specific to these speakers. These results are discussed within the context of the larger study in Chapter 9.

Chapter 5: Mixed Dialect Group (NS_{AE}-NS_{IE}, Interdialectal Condition)

5.1 Introduction

PC has been examined with mixed results in dialectally variant conditions (Kim, Horton, & Bradlow, 2011; Krivokapic, 2013). Comparing rhythmic properties across AE and IE dialects, Krivokapic (2013) found limited indication of convergence after a synchronized reading task. Using a durational measure for rhythm, she found that the timing of one female IE speaker became more stress-timed and the American speakers showed a tendency towards a more syllable-timed rhythm. Exploring interactions between speakers of different language backgrounds, Kim et al. (2011) found that linguistic distance modulates PC during a diapix task. Speakers who were linguistically close or spoke the same dialect of a language, southern AE, were more likely to converge than speakers who were in the intermediate group (they spoke different dialects of the same language) or the far group (they spoke different L1s) conditions. Although they examined dyads that varied in dialects of AE and Korean, Kim et al.'s (2011) study did not explore speakers of different national varieties of English. The current experiment builds on the findings from these two studies by using a spectral measure of rhythm and examining convergence within speakers of two different national varieties of English, namely AE and IE.

The main purpose of this experiment was to examine changes in vowels and rhythm across speakers of different English dialects. Specifically, speakers of AE and IE were selected because these dialects share vowels but are considered rhythmically distinct. Recall that rhythmically, AE is considered stress-timed whereas IE is mixed. In addition, Fuchs (2012) found that IE is characterized by less vocalic and consonantal variation than British English and by extension AE. Importantly, he also noted that L1 background of IE speakers did not affect this finding. That is, regardless of the local

language variety used by the speakers (*e.g*. Hindi, Marathi, etc.), they all produced a consistently similar variety of IE. Furthermore, using the term 'L1' to denote the other Indian language spoken by the subjects is misleading here. Most IE speakers are typically multilingual and proficient speakers of English as well as one or more Indian languages. For the purposes of this study, it would be more appropriate to consider IE speakers to be multilingual, *i.e*. native IE speakers who speak one or more Indian languages, rather than L2 speakers of English.

Recruiting for this group for the current study raised the question of controlling for further subdivisions of dialects within AE and IE speech. Like the native group, dialect as defined by the particular region of the US or India where a participant was raised was allowed to vary. Given the testing location, it was assumed that native AE speakers would speak a variety of General American with some southern features. Controlling for IE regional variation was not feasible given the availability of subjects in the Austin area. Furthermore, no extensive literature, and as a result, no systematic way of differentiating among regional varieties of IE exists currently. A detailed language background questionnaire was administered for each participant and the results are provided in Section 5.3.1.

5.2 HYPOTHESES

5.2.1 Vowels

Literature on speech accommodation suggests that convergence is the default process and divergence takes place when the need to mark oneself as socially distinct from a partner arises (Babel, 2010; Shepard et al., 2001). Since the current study did not contain any explicit social manipulations, it was expected that interactions between

proficient speakers of different national varieties of English would lead to convergence. From the six vowels, the low, back /ɑ/ vowel was pronounced as /ɔ/ by all IE speakers. Thus, this was a marked vowel for this group. The low frequency of words that contained this vowel (in its marked form) would elicit stronger episodic traces (Goldinger, 1998). Thus, this vowel would be the most susceptible to convergence. The other five vowels exist in the vowel inventories of both dialects and were expected to show convergence.

In formants, convergence would be noted by a reduction in the difference in provider and receiver post-task values. Sys-IS and spec-IS would indicate convergence via an increase in post-task values.

5.2.2 Rhythm

Convergence would be characterized by either a decrease in the distance between the speaker pair's centroids or a reduction in the height of both the provider and the receiver's centroids as outlined in Section 3.5.2. Conversely, divergence would be indicated by an increase in the distance of a speaker pair's centroids or an increase in the heights of both speakers' centroids. Since all speakers are proficient speakers of two different national varieties of English, the rhythmic properties of each dialect would be salient to the other speaker and, as per Goldinger's (1998) prediction, also susceptible to convergence. However, it should be noted that analyses of pilot data for this experiment suggested that speakers of IE and AE, regardless of sex, diverge in spectral rhythm (Rao et al., 2011).

As noted in the introduction, IE shows less vowel and consonant reductions and deletions as compared to British English. Thus, AE is expected to show more variable rhythm than IE as indicated by a higher centroid for the AE speakers than the IE

speakers. Rhythmic differences specific to the speaker pair would also be noted by EMS + centroid.

5.3 METHODOLOGY

Methods and stimuli are as described in the chapter on methodology, Chapter 3. Information specific to this experiment is provided below.

5.3.1 Participants

24 speakers (12 male) participated in the study. Their ages ranged from 18-51 years old (mean = 22.47 years (SD = 7.2)). Participants did not have any known speech or hearing impairments at the time of recording. They were either students (graduate or undergraduate) at the University of Texas at Austin or professionals living in the greater Austin area.

Twelve speakers were bilingual IE speakers who spoke one or more other Indian languages and 12 were native AE speakers. Besides English, the native IE speakers also spoke other languages fluently (Hindi, Bengali, Gujarati, Punjabi or Tamil). Eight of the twelve IE speakers were trilingual and 4 were bilingual. Two of the speakers also listed French and Russian as L2 languages spoken. One male IE speaker was born in the US but moved to India at the age of 9. For this study, a speaker was considered a native speaker if he or she had been exposed to a language before the age of 10. Based on this criterion, this speaker was considered a native IE speaker and his data was included in the analysis. Another female speaker was born in India but moved to Saudi Arabia at an unknown age. At the time of participation, speakers had spent anywhere from a month to seven years in the US.

Nine of the 12 AE speakers were from Texas. Two of these nine had spent time in Ohio from 0-5 years and 0-6 years and one in Virginia from 0-3 years before moving to Texas. Two had also split their time between another state and Texas, one had lived in Louisiana from 5-7 years and another had lived in Minnesota from 11-14 years. Three other AE speakers were from California, New York and Georgia and moved here at 30, 22 and 24 years of age respectively. All but three had some experience with a second language (Spanish, French, Hebrew, Danish, American Sign Language, Italian, Arabic or German) via high school or college courses as part of a language requirement but were not fluent in any of these languages as indicated in the background language questionnaire.

Participants were assigned receiver or provider roles upon arrival to the lab. The first pair in this condition was assigned roles randomly and subsequent pairs were assigned the opposite roles to counterbalance the design. In this way, 3 IE speakers were providers (paired with AE receivers) and 3 were receivers (paired with AE providers). Participants took an average of 21.59 minutes (SD = 14.47) to complete the map task. Two male pairs were stopped after completing three of the four maps due to time considerations. These pairs spent a mean time of 47.70 minutes before being stopped.

5.3.2 Rhythm

Two sentences of the 11 total sentences from the recording paragraph (sentences 11 and 8) were dropped for all speakers because they were missing from the recordings for two speakers.

5.4 RESULTS

The following section describes the results of the formant midpoint, IS and rhythm analyses in that order. For each section, the descriptive trends are discussed first followed by the statistical findings separated by sex. Alternative descriptive plots for vowels are also provided in Appendix E. Results from all statistical analyses are provided in Appendix D.

5.4.1 Midpoint formant analyses

5.4.1.1 Female speakers

Figure 5.1 shows average formant values for all six vowels separated by role and task for the mixed dialect group for female speakers, respectively. The plots reveal role-based differences for some vowels. For example, female speakers' F1 and F2 values for /o/, /i/, /u/ appear to be larger for providers than for receivers. F2 values of /e/ suggest a change from pre-task to post-task that suggests an increase in the distance between a provider and receiver F2 means.

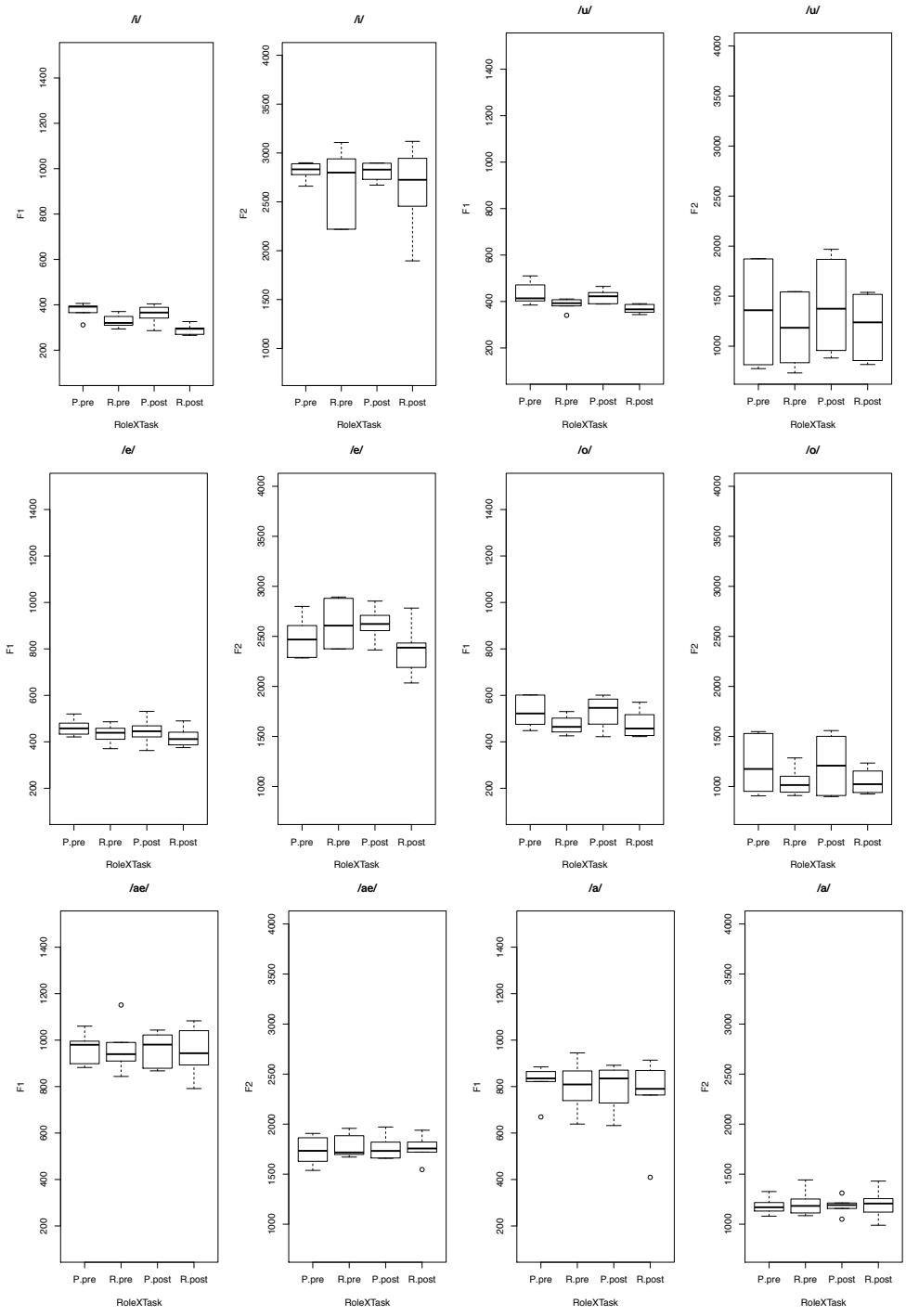


Figure 5.1: Average F1 and F2 values for all female vowels separated by role and task. Vowels are arranged in order of vowel height from top to bottom (/æ/ and /ɑ/ are listed as ‘ae’ and ‘a’ respectively).

Provided below are mean F1 and F2 values for female speakers (Tables 5.1 and 5.2 respectively). Standard errors (SE) are in parentheses. These show that IE speakers regardless of role decreased their F1 values from pre-task to post-task but the AE speakers did not. IE speakers also increased their F2 values. AE speakers who were providers increased their F2 values but the AE speakers who were receivers decreased theirs from pre-task to post-task.

	Providers		**Receivers**	
Dialect	**Pre**	**Post**	**Pre**	**Post**
AE	613.369 (54.68)	618.225 (54.87)	559.757 (59.33)	550.304 (60.65)
IE	581.393 (53.51)	553.878 (54.42)	568.308 (57.87)	536.160 (59.05)

Table 5.1: Mean F1 values for female speakers separated by task, role and dialect.

	Providers		**Receivers**	
Dialect	**Pre**	**Post**	**Pre**	**Post**
AE	1905.636 (137.81)	1922.781 (143.02)	1842.516 (160.63)	1819.268 (155.77)
IE	1686.058 (185.57)	1742.986 (185.85)	1658.313 (181.93)	1588.282 (158.59)

Table 5.2: Mean F2 values for female speakers separated by task, role and dialect.

A 6 X 2 X 2 X 2 mixed-design ANOVA (evaluated with significance level alpha adjusted at 0.025), using F1 as a dependent variable (DV), vowel type (/æ, ɑ, i, u, e, o/)

and task (pre or post) as within-subjects factors and role (receiver or provider) and dialect (AE or IE) as between subjects factors was conducted. The results revealed main effects of vowel ($F(5, 10) = 587.15$, $p < 0.001$) and role ($F(1, 75) = 9.580$, $p < 0.01$). Another mixed design ANOVA with the same IVs but F2 as the DV revealed main effects of vowel ($F(5,10) = 168.92$, $p < 0.001$), dialect ($F(1,72) = 46.382$, $p < 0.001$) and role ($F(1,72) = 8.518$, $p < 0.01$). AE speakers produced all vowels with larger F2 values than IE speakers. Furthermore, providers also produced all vowels with larger F1 and F2 values than receivers suggesting that they produced more fronted and lower vowels than the receivers. There were no other main effects or significant interactions. Significant F1 and F2 values for main effects of dialect and role are shown in Figures 5.2 and 5.3 respectively.

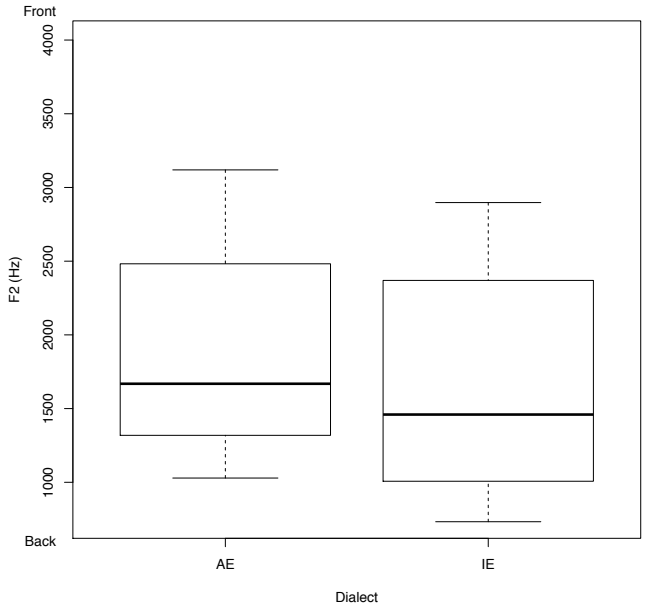


Figure 5.2: F2 means for female speakers separated by dialect. 'Front' and 'back' indicate vowel quality.

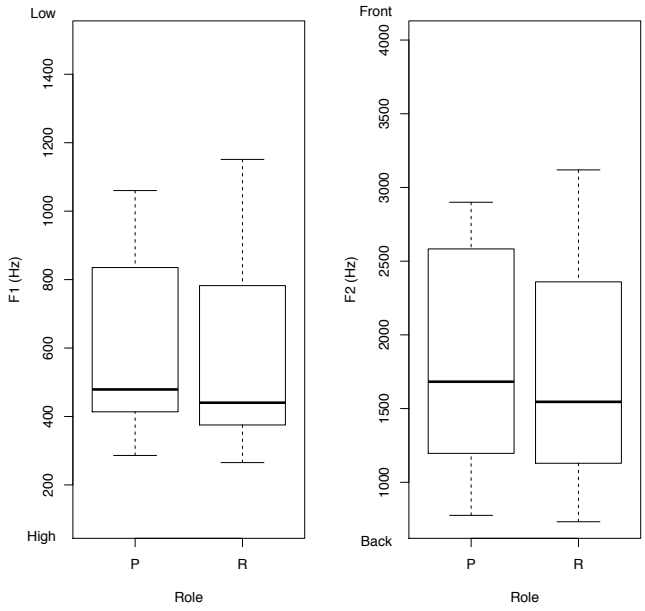


Figure 5.3: F1 and F2 means for female speakers separated by role. 'Low', 'high', 'front' and 'back' indicate vowel quality.

5.4.1.2 Male speakers

Figure 5.4 shows average formant values for all six vowels separated by role and task for the mixed dialect group for male speakers. These plots show that speakers did not appear to alter their F1 or F2 values from pre-task to post-task. The plots also reveal role-based differences for some vowels. For example, receivers' F1 values for /æ/ and /ɑ/ appear to be larger for receivers than for providers.

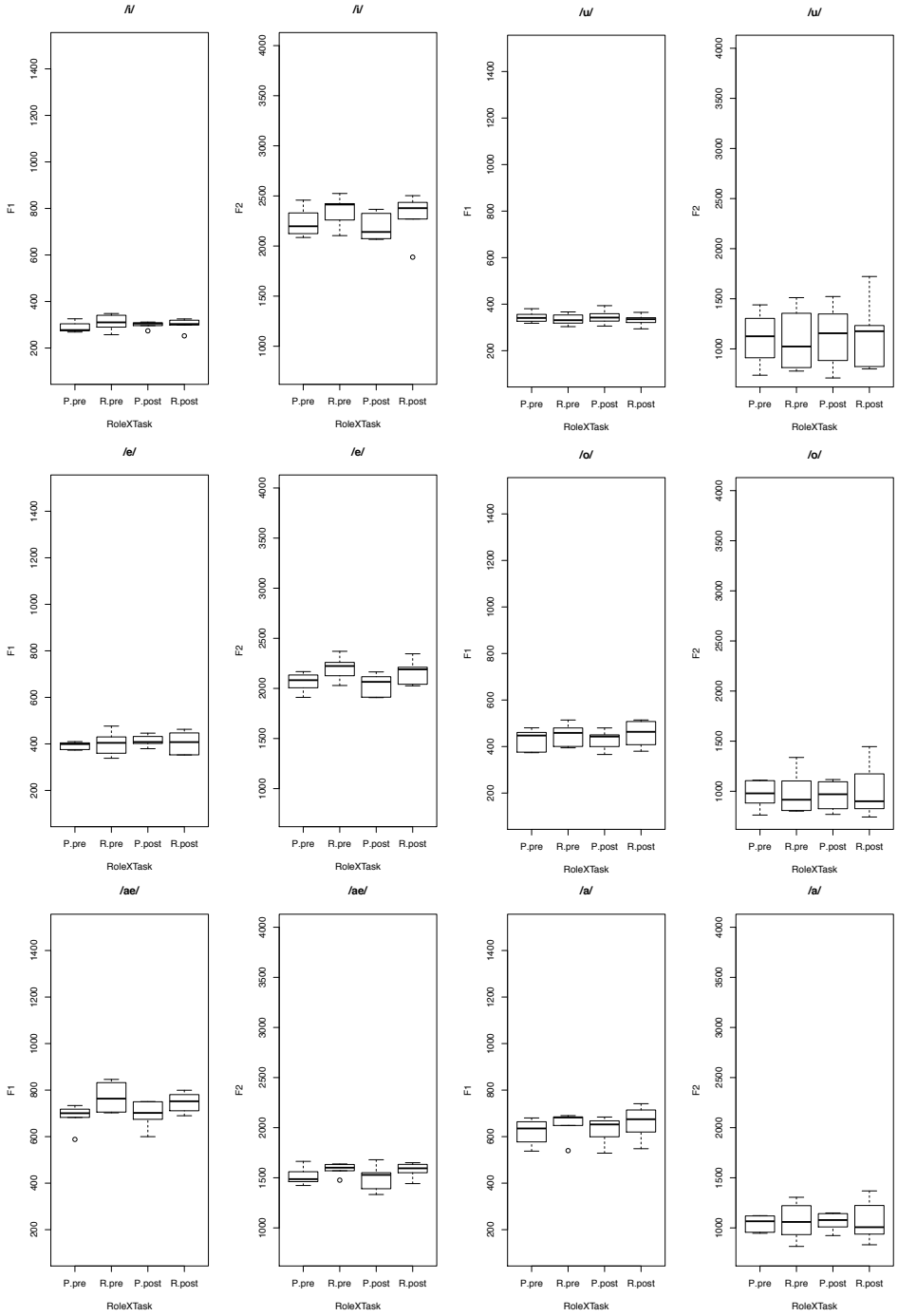


Figure 5.4: Average F1 and F2 values for all male vowels separated by role and task. Vowels are arranged in order of vowel height from top to bottom (/æ/ and /ɑ/ are listed as 'ae' and 'a' respectively).

Provided below are mean F1 and F2 values for male speakers (Tables 5.3 and 5.4 respectively). Standard errors (SE) are in parentheses. These show that providers regardless of dialect decreased their F2 values slightly from pre-task to post-task. Providers also increased their F1 values from pre-task to post-task. The receivers showed dialectal differences in F1, AE receivers increased their F1 values whereas IE receivers decreased theirs from pre-task to post-task.

	Providers		**Receivers**	
Dialect	**Pre**	**Post**	**Pre**	**Post**
AE	476.730 (39.08)	483.927 (38.43)	476.359 (41.12)	482.490 (42.16)
IE	444.919 (32.62)	454.545 (33.72)	496.985 (43.20)	485.266 (41.03)

Table 5.3: Mean F1 values for male speakers separated by task, role and dialect.

	Providers		**Receivers**	
Dialect	**Pre**	**Post**	**Pre**	**Post**
AE	1548.376 (108.40)	1539.229 (104.37)	1669.547 (129.52)	1675.802 (125.86)
IE	1429.280 (138.92)	1419.848 (135.69)	1423.432 (145.55)	1414.279 (139.20)

Table 5.4: Mean F2 values for male speakers separated by task, role and dialect.

A 6 X 2 X 2 X 2 mixed-design ANOVA (with significance level alpha adjusted at 0.025), using F1 as a DV, vowel type (/æ, ɑ, i, u, e, o/) and task (pre or post) as within subjects factors and role (receiver or provider) and dialect (AE or IE) as between subjects factors revealed main effects of vowel ($F(5, 10) = 494.85$, $p < 0.01$) and role ($F(1,72) =$

9.708, $p < 0.001$) and a three-way interaction between vowel, dialect and role ($F(5, 72) = 2.852$, $p < 0.05$). A separate mixed-design ANOVA with the same IVs as the F1 model but F2 as a DV revealed main effects of vowel ($F(5, 10) = 210.61$, $p < 0.001$), dialect ($F(1, 72) = 72.433$, $p < 0.001$) and role ($F(1, 72) = 7.895$, $p < 0.01$) and two two-way interactions: vowel x dialect ($F(5, 72) = 8.930$, $p < 0.001$) and dialect x role ($F(1, 72) = 9.426$, $p < 0.01$). Post-hoc tests of simple effects using $\alpha = 0.002$ revealed that F1 values were significantly larger for /æ/ of IE receivers than IE providers (Figure 5.5), *i.e.* IE receivers produced lower /æ/ vowels than IE providers.

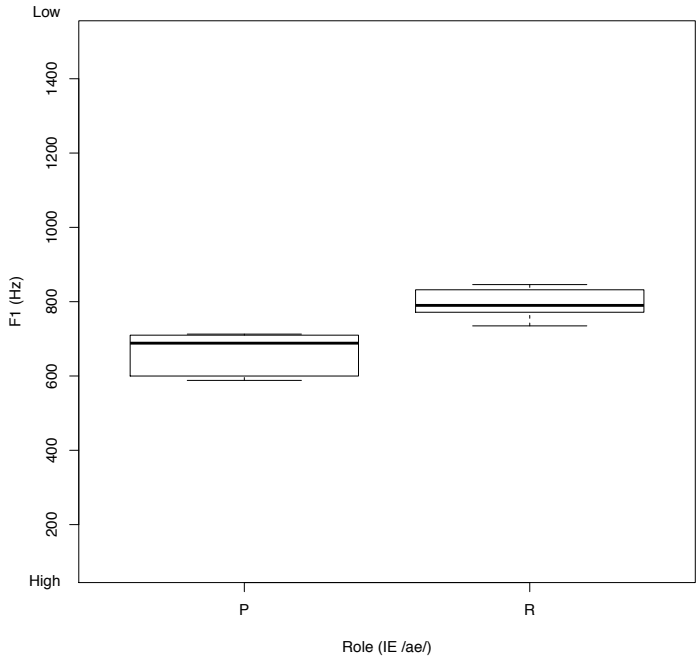


Figure 5.5: Male F1 means for IE /æ/ separated by role. 'Low' and 'high' indicate vowel quality.

For /ɑ/, F1 values were significantly larger for AE speakers than for IE speakers (Figure 5.6), *i.e.* AE speakers produced lower /ɑ/ vowels than IE speakers.

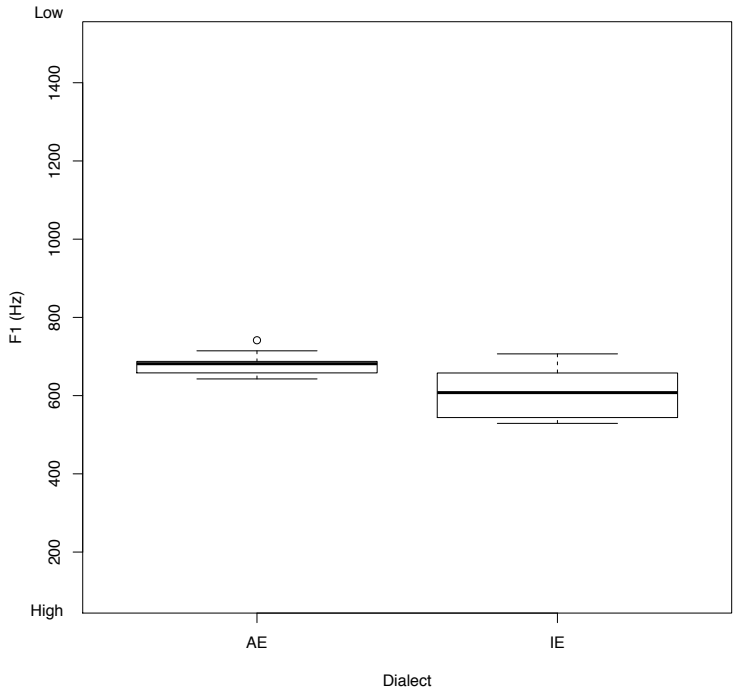


Figure 5.6: Male F1 means for /ɑ/ separated by dialect. 'Low' and 'high' indicate vowel quality.

Similar post-hoc tests also revealed significantly larger F2 values of AE speakers than IE speakers for /ɑ/, /u/ and /o/, *i.e.* AE speakers produced more fronted /ɑ/, /u/ and /o/ than the IE speakers. These significant differences are depicted in Figure 5.7.

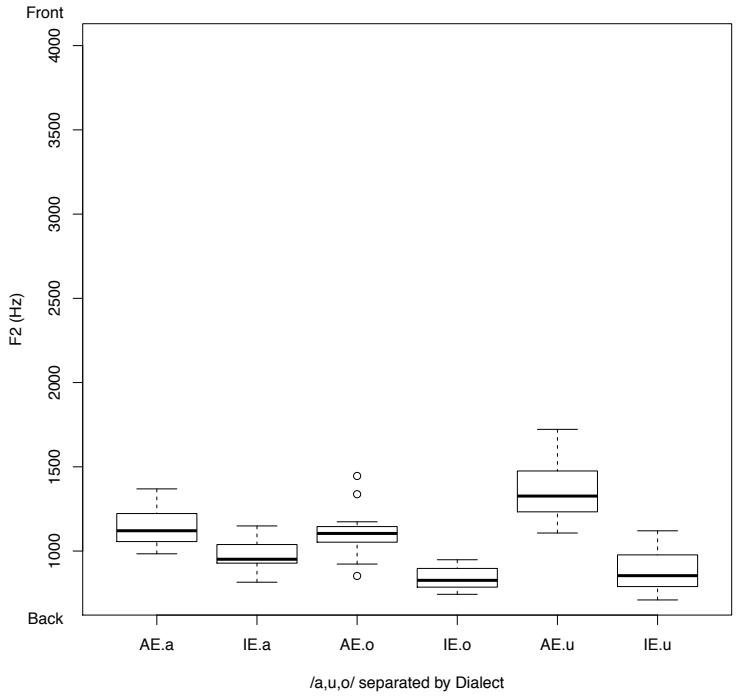


Figure 5.7: Male F2 means for significant vowels separated by dialect. AE /ɑ, u and o/ were more fronted than IE /ɑ, u and o/ (plot shows /ɑ/ as /a/). 'Front' and 'back' indicate vowel quality.

Figure 5.8 shows F2 differences based on role. Providers had smaller F1 and F2 values, *i.e.* providers produced vowels that were higher and further back than the receivers.

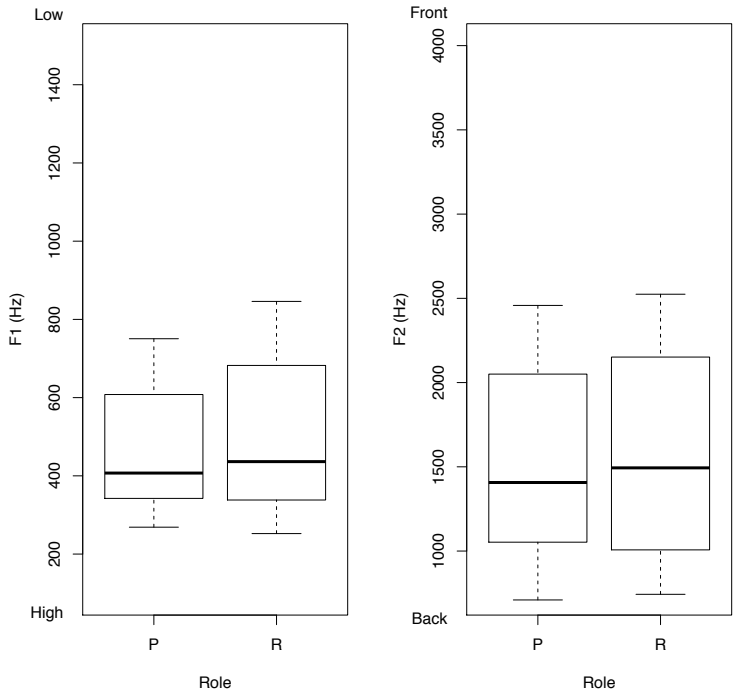


Figure 5.8: F1 and F2 means for male speakers separated by role. 'Low', 'high', 'front' and 'back' indicate vowel quality.

Figure 5.9 shows F2 differences based on dialect. AE speakers produced vowels with larger F2 values, *i.e*. AE speakers produced more fronted vowels as compared to the IE speakers.

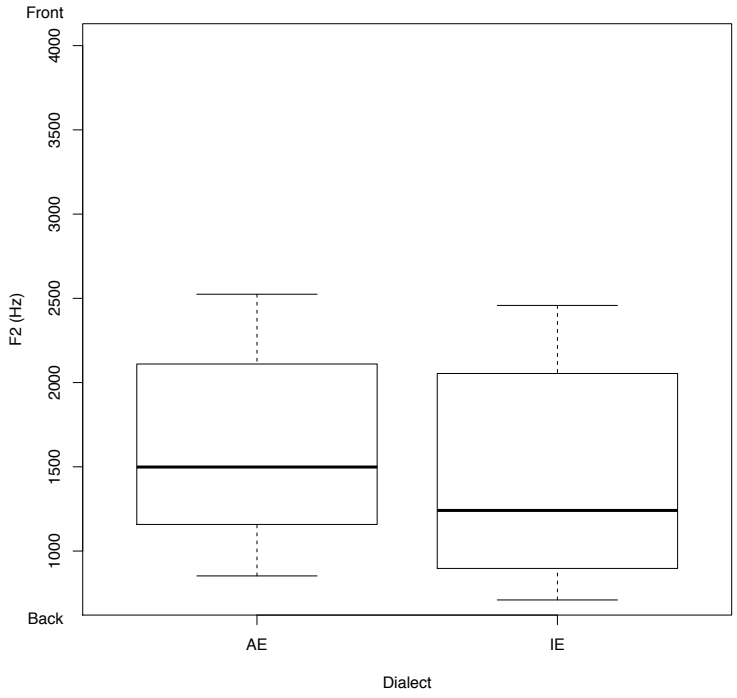


Figure 5.9: F2 means for male speakers separated by dialect. 'Front' and 'back' indicate vowel quality.

5.4.2 Interlocutor Similarity (IS)

The mean values for all ISs for both male and female speakers are provided in Table 5.5. An increase in similarity scores for both male and female IS demonstrates a decrease in distance between the vowel midpoints of each dyad's receiver and provider for both men and women or convergence. Analogously, a decrease in the similarity scores denotes divergence. It can be seen from this table that for female speakers, both systemic and specific IS increased after the. Male sys-IS values decreased from pre-task to post-task.

	Pre-task	**Post-task**
Sys-IS (women)	0.9787 (0.00)	0.9833 (0.00)
Spec-IS (women)	0.9939 (0.00)	0.9959 (0.00)
Sys-IS (men)	0.9900 (0.00)	0.9872 (0.00)
Spec-IS (men)	0.9971 0.00)	0.9973 (0.00)

Table 5.5: Mean IS separated by task for male and female speakers (SE in parenthesis are 0 up to two significant digits).

Two bootstrap analyses of repeated-measures ANOVAs with sys-IS as the DV and task as the IV were run on male and female data separately. Sys-IS and spec-IS are a measure of similarity between the receiver and provider. As a result, role and dialect cannot be tracked anymore. These bootstraps revealed significant effect of task for both male and female speakers. However these effects were small as noted by the 97.5% CI being very close to 1 in Figure 5.10.

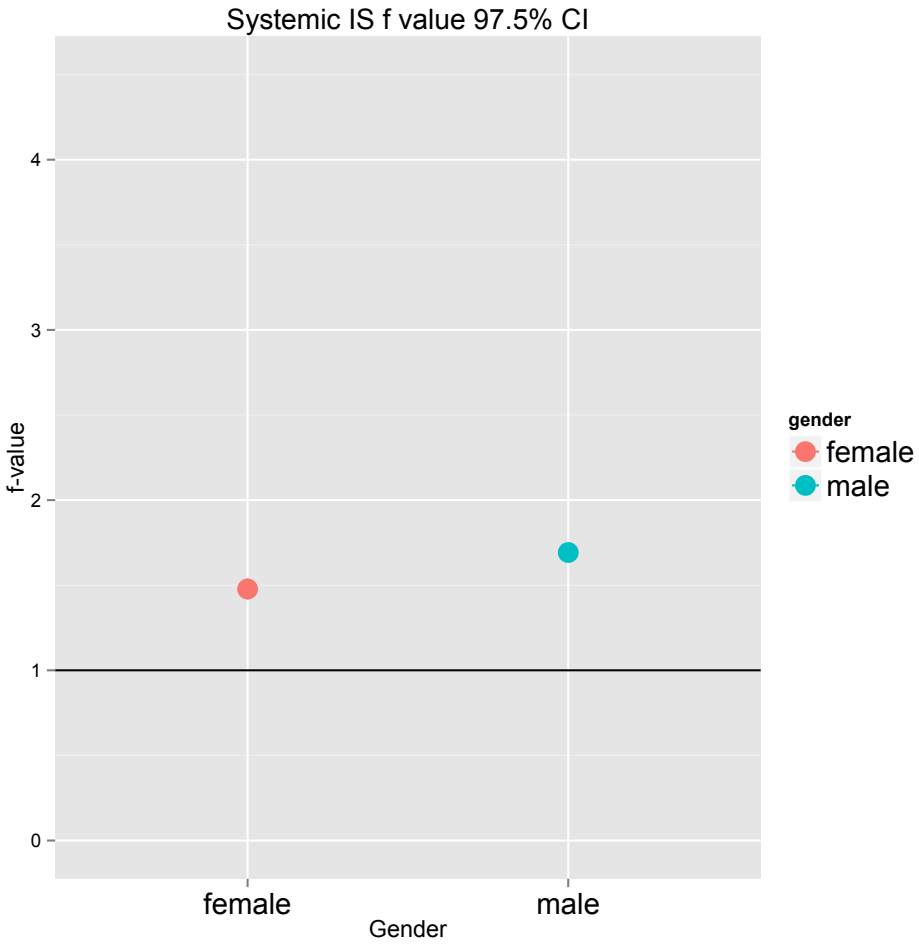


Figure 5.10: 97.5% CI of f-values for sys-IS (male and female speakers). Red points indicate female and blue indicate male f-values.

Two analogous bootstraps with spec-IS as DV and task and vowel as IVs showed a main effect of task and an interaction between vowel and task for the female speakers but not the male speakers. 97.5% CIs for spec-IS are provided in Figure 5.11. Means for spec-IS values separated by task and vowel are provided below (Figure 5.12).

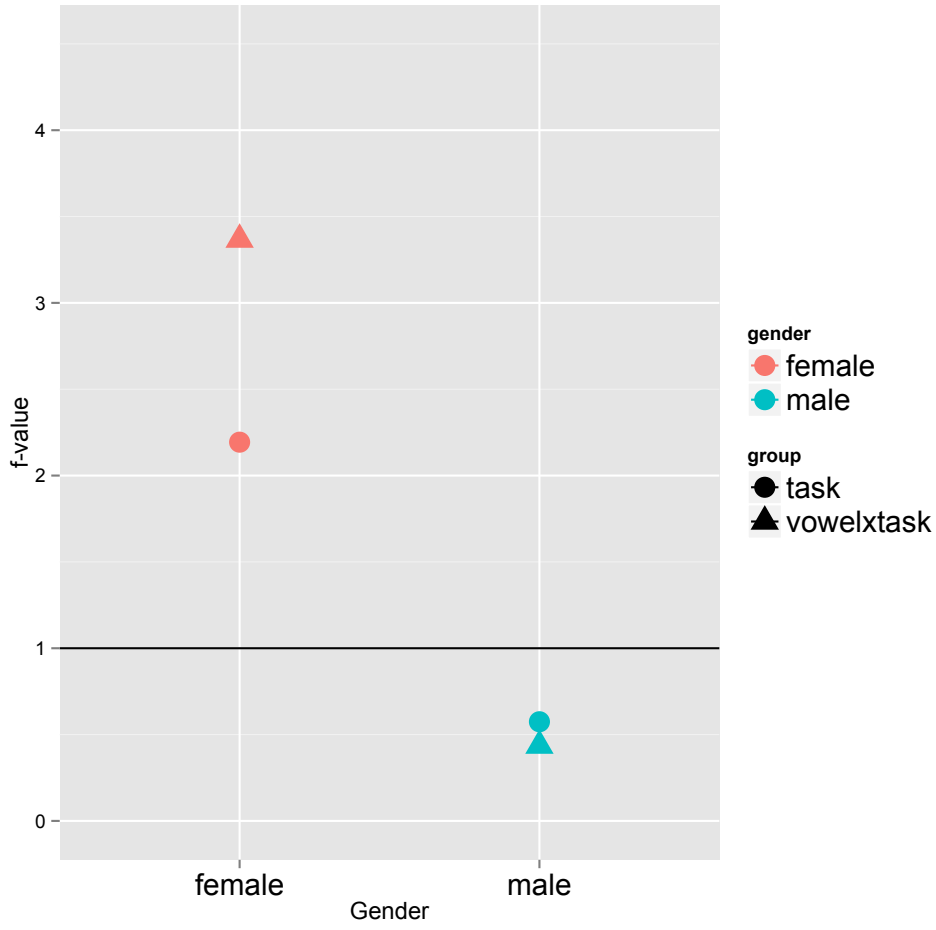


Figure 5.11: 97.5% CI of f-values for spec-IS. Red points indicate female and blue indicate male f-values. Circles indicate task f-values whereas triangles indicate f-values for the vowel X task interaction.

Figure 5.12 shows the spec-IS for each vowel before and after the interaction for female speakers. Arrows in the plot indicate an increase or decrease in similarity after map task completion. Arrows pointing up denote convergence whereas arrows pointing down denote divergence. After the interaction, /u/ and /o/ show increased similarity, /ɑ/ and /æ/ show decreased similarity and /e/ and /i/ showed no change.

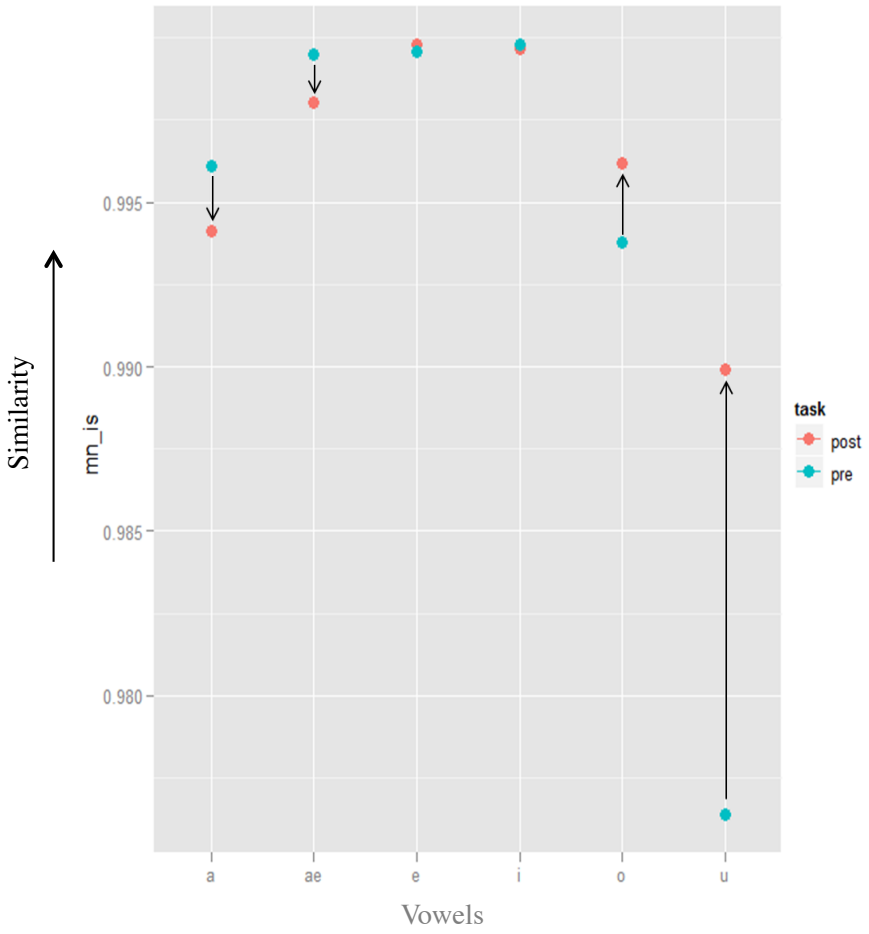


Figure 5.12: Plot of the mean spec-IS of each vowel for female speakers. Arrows indicate increase or decrease in similarity (/ɑ/ and /æ/ are listed as 'a' and 'ae' respectively in figure). Red points indicate post-task and blue indicate pre-task means for each vowel.

5.4.3 Rhythm

5.4.3.1 Female speakers

All female centroid averages are provided in Table 5.6. This table shows that female dyads regardless of dialect had lower post-task centroids compared to pre-task

centroids for both providers and receivers suggesting less variable rhythm after the interaction.

	Providers		**Receivers**	
	Pre-task	**Post-task**	**Pre-task**	**Post-task**
AE	9.006 (0.25)	8.848 (0.31)	9.108 (0.43)	9.052 (0.26)
IE	8.188 (0.37)	7.825 (0.35)	8.912 (0.35)	8.331 (0.36)

Table 5.6: Mean rhythm centroids for female speakers separated by task, role and dialect.

A 2 X 3 X 2 X 2 nested factorial ANOVA (with significance level alpha adjusted at 0.025) using role (receiver or provider), speaker pair (speaker pair 1-3), task (pre or post) and dialect (AE or IE) as IVs and EMS + centroid values as the DV was run on the female data. Because native language as a factor was also included in this analysis, speaker pairs had to be coded in a manner that collapsed two speaker pairs into one. Thus, each speaker pair contained one $Provider_{IE}$ - $Reciever_{AE}$ and one $Provider_{AE}$-$Reciever_{IE}$ combination.

The analysis of female centroids revealed main effects of speaker pair ($F(2, 128) = 7.742$, $p < 0.001$) and dialect ($F(1, 8) = 15.13$, $p < 0.01$). Two interactions: role x speaker pair x dialect ($F(2, 128) = 6.153$, $p < 0.01$) and task x speaker pair x role ($F(2, 128) = 4.017$, $p < 0.05$) were also noted. No other main effects or interactions were significant.

The mean centroid value for women who spoke AE was 9.004 (SE = 0.16) and for women who spoke IE was 8.314 (SE = 0.18). Post-hoc Bonferroni tests at $\alpha = 0.008$ (alpha adjusted) revealed that speaker pairs 1 and 2 had significantly different centroids from each other, such that speaker pairs 1 (mean = 8.232, SE = 0.20) had lower centroids

than speaker pairs 2 (mean = 9.135, SE = 0.25). Other post-hoc tests at $\alpha = 0.008$ significance revealed significantly different centroids for speaker pairs 2 based on dialect, role and task. For these pairs, AE speakers (mean = 9.977, SE = 0.25) had higher centroids than IE speakers (mean = 8.293, SE = 0.38) and receivers (mean = 9.700, SE = 0.37) had higher centroids than providers (mean = 8.570, SE = 0.31). Furthermore, pre-task centroids (mean = 9.829, SE = 0.30) were higher than post-task centroids (mean = 8.441, SE = 0.36) for this group. Lastly, AE speakers from speaker pairs 3 also showed an effect of role such that AE receivers (mean = 8.221, SE = 0.16) had higher centroids than AE providers (mean = 9.58, SE = 0.33). Higher centroids indicate a more variable rhythm, suggesting that female receivers showed more variable rhythm than female providers. Furthermore, for speaker pairs 2, rhythmic variability was higher pre-task than post-task, and for speaker pairs 3, AE receivers had more rhythmic variability than AE providers.

5.4.3.1 Male speakers

All male centroid averages are provided in Table 5.7. This table shows that male pairs raised their centroids regardless of task from pre-task to post-task indicating a more variable rhythm.

	Providers		**Receivers**	
	Pre-task	**Post-task**	**Pre-task**	**Post-task**
AE	7.648 (0.40)	8.186 (0.29)	8.599 (0.38)	8.761 (0.25)
IE	8.596 (0.32)	8.793 (0.32)	8.957 (0.32)	9.454 (0.30)

Table 5.7: Mean rhythm centroids for male speakers separated by task, role and dialect.

A nested factorial ANOVA of male data using the same IVs and DV revealed a main effect of role ($F(1, 8) = 39.277$, $p < 0.001$), dialect ($F(1, 96) = 11.758$, $p < 0.001$) and task ($F(1, 48) = 4.4658$, $p < 0.05$) and an interaction between speaker pairs and role ($F(5, 32) = 14.8951$, $p < 0.001$). No other main effects or interactions were significant.

Male receivers (mean = 8.943, SE = 0.16) had higher centroids than male providers (mean = 8.306, SE = 0.17). Male IE speakers (mean = 8.950, SE = 0.16) had higher centroids than male AE speakers (mean = 8.299, SE = 0.17). Lastly, post-task centroids (mean = 8.799, SE = 0.15) for male speakers were higher than for pre-task items (mean = 8.450, SE = 0.18). Higher post-task centroids indicated an increase in rhythmic variability in men after the task. Post-hoc tests of simple effects at 0.004 significance level revealed significantly different centroids for male speaker pairs 3's receivers and providers. The receivers (mean = 9.250, SE = 0.25) of this group displayed higher centroids than the providers (mean = 7.494, SE = 0.28).

5.5 DISCUSSION

5.5.1 Vowels

Midpoint vowel formant analyses noted a number of dialectal differences. Female speakers of AE exhibited higher F1 and F2 values than speakers of IE, indicating that AE speakers produced more fronted and lowered vowels than IE speakers. This was noted particularly in the AE /ɑ/ and /u/ vowels which were more fronted than the IE versions for female speakers.

Male speakers of AE also exhibited higher F1 and F2 values than speakers of IE, indicating that AE speakers produced more fronted and lowered vowels than IE speakers. In addition to these vowels, male AE realizations of /o/ were more fronted and /æ/ were

lowered compared to the IE versions. Though IE vowel acoustics are less studied, the more fronted versions of the AE /ɑ/, /u/ and /o/ vowels are as expected (Labov, 2006).

Midpoint analyses also revealed other vowel, dialect and role specific differences. Female providers produced more fronted and lowered vowels than female receivers regardless of dialect. Female providers in this condition also fronted their /u/ more than the receivers did. Additionally, female AE providers fronted their /o/ more than female AE receivers. Male receivers produced more fronted and lowered vowels than male providers. Within this set, male IE receivers produced lower vowels than male IE providers. These speakers set up acoustic distinctions based on the role they were assigned in the beginning of the task. Reasons why such role-based differences are noted are unclear.

Even though the midpoint formant analysis did not show an effect of task on vowel changes, the sys-IS and spec-IS scores revealed some vowel adaptations due to the task. The sys-IS scores increased suggesting that female speakers increased in vowel similarity signaling vowel convergence. The spec-IS scores further specified convergence in /u/ and /o/ but divergence in /ɑ/ and /æ/. /i/ and /e/ showed no substantial change.

For the purposes of this study, /ɑ/ is of particular interest because it is a marked vowel across these dialects. For IE speakers, this vowel was pronounced as /ɔ/ (*e.g.* hawd) and as per spec-IS, it diverged for AE and IE women. Examining the mean formants reveals that both AE and IE female speakers maintained their F2 values but lowered their F1 values resulting in raised /ɑ/ or /ɔ/ vowels (Figure 5.13).

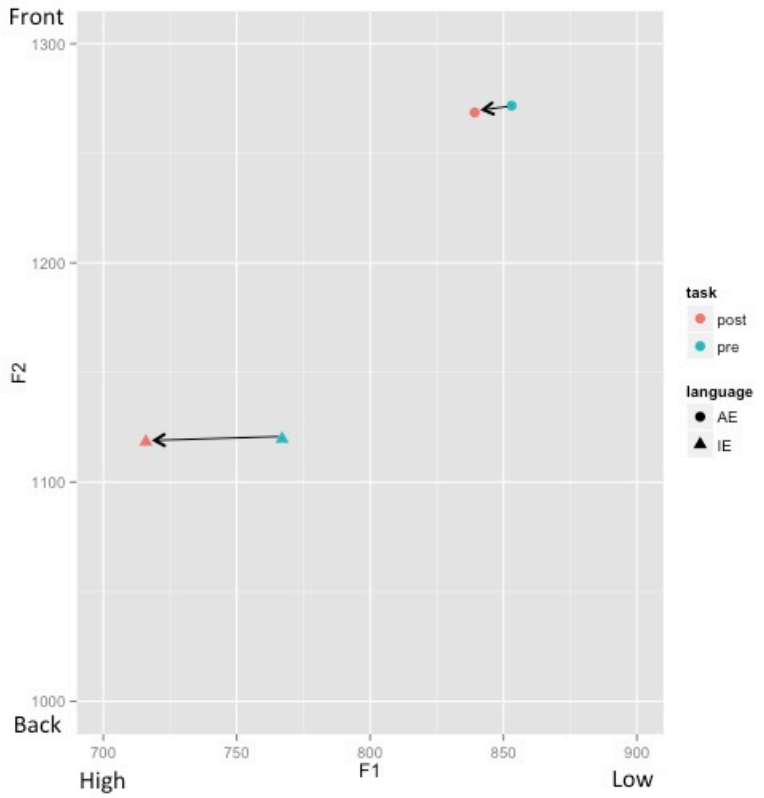


Figure 5.13: Mean F1 and F2 for /ɑ/ separated by task and dialect (female speakers).

/ɑ/ and /ɔ/ vary mostly along the F1 dimension, with /ɑ/ having a higher F1 than /ɔ/. While AE women raised their /ɑ/ vowels to make them more /ɔ/-like, presumably to make them more similar to the IE vowel, IE women made their /ɔ/ vowels more /o/-like by raising them further. Thus although the AE speakers seem to have moved their vowel toward their interlocutor's vowel category (*i.e.* converged), IE speakers moved their vowel to a greater extent in the direction that indicates divergence. Spec-IS values for this vowel indeed decreased from pre-task to post-task demonstrating divergence for this vowel.

Another marked vowel for this group is /u/ which is produced highly fronted by most AE speakers and is fronted to a greater extent by southern varieties of AE such as Texas (Labov, Ash, & Boberg, 2006). Even though AE sub-dialects were not controlled, most of the AE speakers were from Texas. Moreover, Austin has been noted as a city that produces extremely fronted /u/ vowels (Labov et al., 2006) resulting in all participants having substantial exposure to an AE dialect with fronted /u/ vowels. Unlike /ɑ/, spec-IS detected convergence in female /u/ vowels. The mean formants of this vowel before and after the task (Figure 5.14) suggest that IE speakers were responsible for the adaptation. They decreased their F1 and increased their F2 values to create more raised and fronted /u/ vowels. Thus, female IE speakers fronted their /u/ vowels with respect to their AE partners resulting in PC.

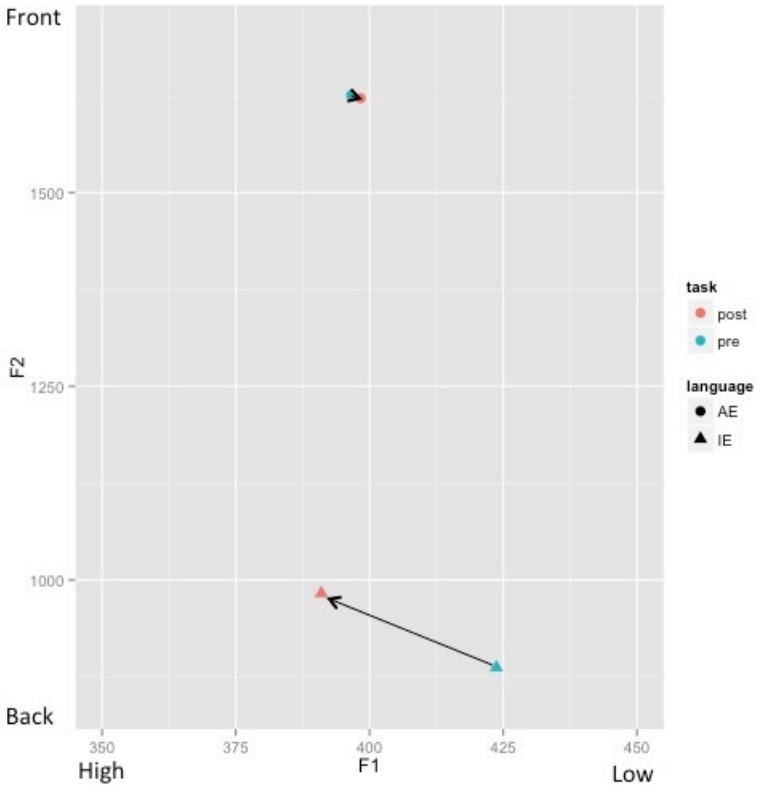


Figure 5.14: Mean F1 and F2 for /u/ separated by task and dialect (female speakers).

According to spec-IS scores, two more vowels showed task specific changes. /o/ converged and /æ/ diverged for the female speakers. Labov et al. (2006) stated that /o/ is produced in a fronted manner by AE speakers even though that fronting is slight compared with /u/. Figure 5.15 shows that both AE and IE speakers altered their F2 values to converge in their /o/ vowels, albeit these movements were small. AE speakers backed their /o/ vowels whereas IE speakers fronted theirs to converge. In addition to altering their /o/ vowels in the F2 dimension, speakers also altered them in the F1 dimensions: AE speakers lowered their /o/ vowels while IE speaker raised their /o/

vowels. However, F1 for /o/ is not marked for these dialects. Convergence via spec-IS was noted in the dimension these vowels are marked for these dialects.

PC in /u/ and /o/ is particularly interesting because both vowels tend to be more back in IE than in AE (Wiltshire & Harnsberger, 2006). The convergence noted in these vowels suggests that both AE and IE speakers were capable of altering their formant values to adapt to their partner's speech.

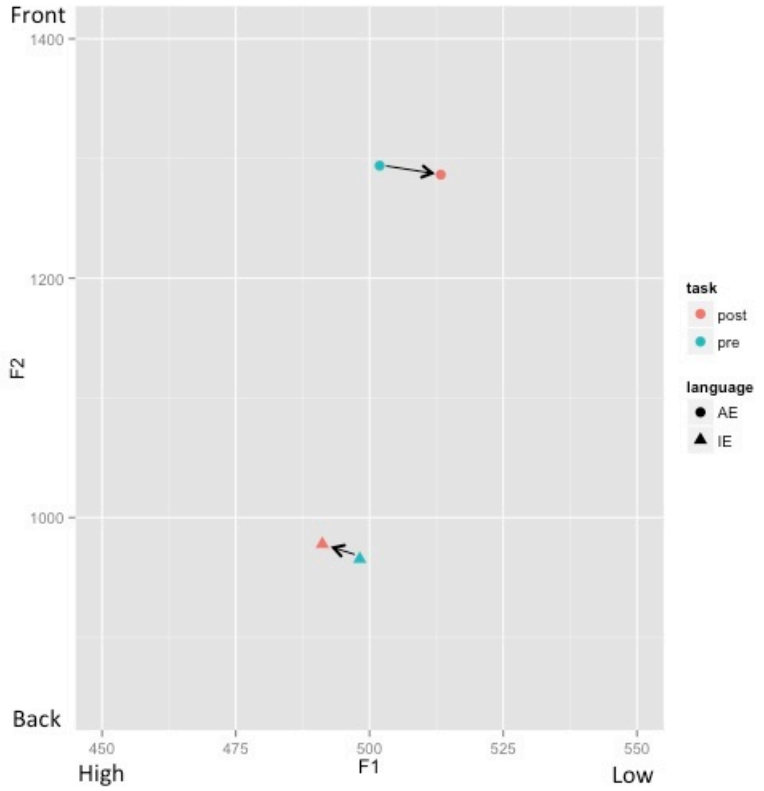


Figure 5.15: Mean F1 and F2 for /o/ separated by task and dialect (female speakers).

Unexpectedly, divergence was noted in /æ/. Figure 5.16 shows F1 and F2 values for /æ/ for women separated by dialect before and after the map task. It can be seen that the main source of divergence in this vowel was due to IE speakers raising their /æ/ vowels. The reasons for this divergence are unclear. This vowel is present in both dialects and the F1 and F2 averages are not very different for the IE and AE versions.

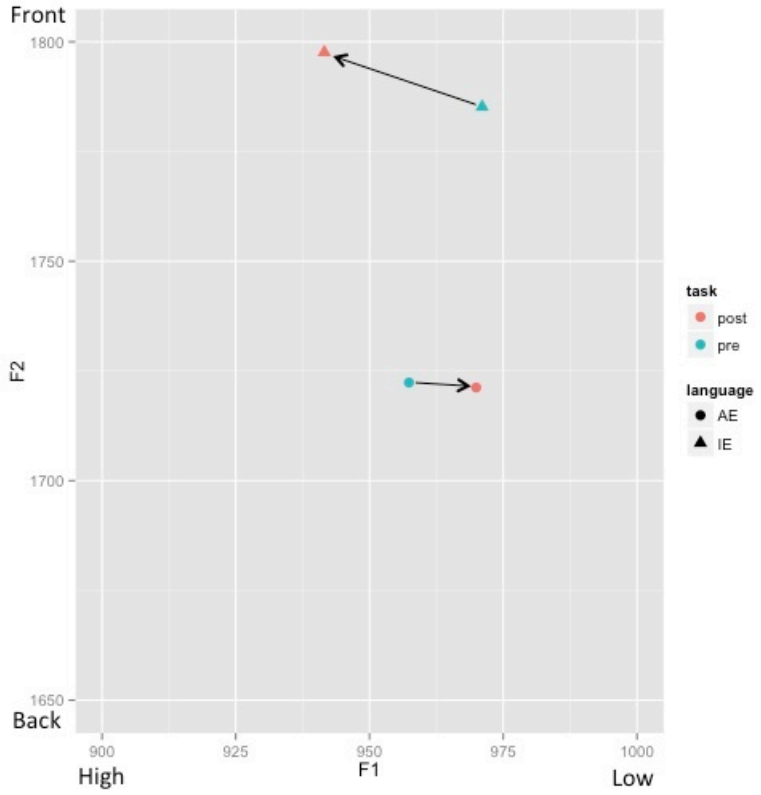


Figure 5.16: Mean F1 and F2 for /æ/ separated by task and dialect (female speakers).

In men, sys-IS scores showed a slight drop from pre- to post-task suggesting divergence.

The predictions made above were that both men and women would show the same direction of adaptation. However, it was not expected that men and women would show opposing adaptations. In women, convergence targeted rounded vowels that were present in both dialects' vowel inventories but are produced in a marked manner: fronted in AE or backed in IE. In contrast, divergence targeted /ɑ/ (or /ɔ/) and /æ/. Based on Goldinger (1998), the prediction had been that /ɑ/ (or /ɔ/) would be more susceptible to convergence because it is unique to IE and would create stronger imitation. Divergence suggests that this distinctiveness was important to the speakers and needed to be preserved. With the exception of /æ/, vowel dialectal markers that were acoustically distinct but phonetically similar converged whereas those that were acoustically and phonetically distinct in IE and AE diverged. Men showed a small change indicating systemic divergence in vowels.

Studies such as Babel (2010), which examined speech adaptations in differing national dialects of a language (Australian English or AuE and New Zealand English or NZE), had a specific goal to examine how a NZE speaker's attitude towards their AuE partner affected PC. The vowels in KIT, TRAP and DRESS are different in AuE and NZE. Even though the difference in the DRESS vowel is least salient to NZE speakers, it converged to the greatest extent. Babel suggests that because speakers were aware of the dialectal difference, KIT and TRAP converged to a lesser extent. The current study did not feature the same attitude manipulation that Babel's did. It is possible that speakers with differing dialects are more likely to converge in marked vowels as long as there is a specific reason to do so (*e.g.* to signal agreement). In the absence of such a goal, speakers who do not share the same national dialect may diverge with respect to marked vowels to maintain social distance.

5.5.2 Rhythm

All male receivers demonstrated a higher centroid than providers, suggesting that providers showed less variable rhythm than receivers. This is a similar trend to that already noted for the pairs in the native language group (Chapter 4). This role difference likely arises from the fact that receivers tended to ask more questions, make clarifications and possibly had more dysfluencies and hesitations than the providers. Such variable speech would lead to an increase in the intensity of the higher EMS frequencies leading to a higher centroid compared to simple statements. If receivers were asking more questions than the providers, it is possible that their rhythm would be more variable than that of the providers. It is also possible that providers switched to clear speech to ensure task success. This would lead to less variability in the rhythm of providers' speech and, as a result, a lower centroid (Rao & Smiljanic, 2011a). Importantly, the role-based difference in centroids was present in both dialects. Thus, the mixed dialect environment did not affect role-based rhythmic differences.

Two female dyads also showed the same pattern noted in the male dyads. In these pairs, the receivers maintained a lower centroid that the providers. Individual rhythm differences were only noted in four of the six pairs of female speakers.

As mentioned above, adaptations in rhythm would be measured in two ways, either with a decrease in the distance between a dyad's centroids or with an increase or decrease in the height of both the speaker's centroids. Divergence would be marked by either an increase in the distance between a dyad's centroids or retention of the distance between a provider-receiver's centroids but an increase in both their heights. This increase in heights signals an increase in overall energy in the higher frequencies of the amplitude envelope. Conversely, a lowering in height of the centroids would indicate a decrease in rhythmic variability or a switch to clear speech by both speakers. Rhythm

adaptation was noted in men whose post-task centroids were higher than pre-task ones indicating divergence. Rhythm adaptations were also noted in women but they were more speaker-specific. Two female speaker pairs lowered their centroids after the interaction indicating convergence.

Though rhythm convergence has not been studied extensively, Krivokapic (2013) reported rhythm convergence in one pair of female speakers out of four. Specifically, the IE speaker converged to the AE speaker. The current experiment supports Krivokapic's (2013) findings by showing limited evidence for rhythm convergence in women (two out of six dyads converged). However, unlike the men in her study who did not show any change, men in the current study exhibited divergence. Consistent with the reasoning in the vowels section above, it is possible that male speakers were diverging to maintain social distinction from their partners. Two female dyads that converged do not fit this pattern. However, Namy et al (2002) have suggested that women may be paying more attention to indexical features during interactive tasks. It is possible that specific speaker characteristics in these dyads contributed to convergence though additional research is required to determine which indexical features were influences. It should be noted that these pairs did not include the IE speaker from Saudi Arabia.

5.5.3 General trends

Generally, both men and women showed vowel and rhythm adaptations in this group. Men showed divergence in both vowels and rhythm whereas women showed vowel-specific convergence and divergence. Lastly, two female dyads also showed convergence in rhythm.

Convergence is often considered the default adaptation whereas divergence may be a marker of group membership (Bourhis & Giles, 1977; Babel 2010). Though either automatic or social theories can explain convergence, divergence appeals to a social interpretation (Shepard et al., 2001). In the current experiment, if PC was a result of automatic or social processes alone, consistent results would have been noted in both vowels and rhythm. Instead, men and women showed vowel and rhythm adaptations to varying extents as noted in Sections 5.5.1 and 5.5.2 above. These differences suggest that reasons to converge or diverge involve a combination of cognitive as well as social considerations. Possible reasons for these trends are discussed further in the general discussion (Chapter 9).

Given previous phonological descriptions that AE rhythm is different from IE rhythm, it was expected that AE tokens would demonstrate higher centroids than IE tokens. In women, AE rhythm had higher centroids than IE rhythm whereas in men, the exact opposite was noted: IE centroids were higher than AE centroids. Rao & Smiljanic (2011) found that EMS + centroid was capable of separating languages based on rhythm but their findings were inconclusive on EMS + centroid distinguishing between dialects. The current finding that male centroid data patterns in opposition to female centroid data with regard to dialectal differences suggests one of two possibilities. One possibility is that EMS + centroid is not suited to separate rhythm of dialects of the same language. Another possibility is that this particular set of speakers does not clearly exhibit dialectal differences as previously described (Fuchs, 2012, Krivokapic, 2013). These possibilities raise the larger question of whether the rhythmic measure employed here is not sensitive enough to detect a dialectal rhythmic distinction or that rhythmic distinction itself deserves deeper consideration and revision. This question is explored further in the general discussion section to include findings from the other datasets (Chapter 9).

Chapter 6: Mixed Language Group (NS_{AE}-NN_{SP}, Far Condition)

6.1 Introduction

The main purpose of the current experiment was to examine PC in the vowels and rhythm of speakers who do not speak the same language natively. With that purpose in mind, native (L1) speakers of AE were paired with non-native speakers (L2) of English who were native speakers of Spanish. Spanish was selected for its vowels as well as its rhythmic properties.

Only two studies to date have examined PC between L1 and L2 speakers of a language. Kim et al. (2011) found that for dyads comprised of native (AE) and non-native (Korean or Chinese) speakers, divergence or maintenance of pronunciation patterns were equally likely. Lewandowski's (2012) study also reported similar results when comparing the amplitude envelopes of native and non-native (native speakers of German) speakers of English. The amplitude envelope served as a measure of global word-level energy. Like Kim et al. (2011), she found that highly proficient non-native speakers of English converged with their native English-speaking partners whereas the least proficient non-native speakers either diverged or showed maintenance.

Of the six AE vowels used in this experiment, four are part of the Spanish vowel inventory: /i/, /e/, /o/ and /u/. Spanish lacks the back, low vowel, /ɑ/ and the front, low vowel, /æ/. Instead it has the central, low /a/ (Ladefoged, 2001). With regard to the rhythm, Spanish is considered syllable-timed whereas AE is considered stress-timed (Ramus et al., 1999; Dauer, 1983). Perceptual tests on bleached speech, where language-specific segmental information is removed, demonstrated that listeners were capable of distinguishing between English and Spanish, suggesting their reliance on suprasegmental properties, *i.e*. rhythm and intonation (Ramus et al. 2003). Spanish and AE were thus chosen to investigate PC across L1 and L2 in vowels and rhythm given their differences

in both segmental and suprasegmental properties. Even though previous research into regional variation in Spanish found some differences in rhythmic properties (using traditional metrics such as nPVI and varcoV) between two Peruvian varieties of Spanish, from Cuzco and Lima, these dialects were still classified as syllable-timed (O'Rourke, 2008b). Thus, including various Spanish dialects was not expected to impact convergence patterns significantly.

6.2 HYPOTHESES

6.2.1 Vowels

It was expected that native AE and native SP speakers would converge in vowels. Lewandowski (2012) found that convergence was moderated by proficiency. For the purposes of the current study, it was ensured that proficiency scores for L1 Spanish speakers were in the high-mid range proficiency (via the pre-interview questions and the LEAP-Q) in speaking English. Based on Goldinger's (1998) exemplar model, the vowels that are missing from the vowel inventory of Spanish speakers, /ɑ/ and /æ/, would be considered marked for them. Vice-versa, the approximations of these vowels by Spanish speakers would be marked for AE speakers. These vowels were expected to be most susceptible to PC in the speech of both L1 Spanish and L1 AE speakers.

In formants, convergence would be noted by a reduction in the difference in provider and receiver post-task values. Sys-IS and spec-IS would indicate convergence via an increase from pre-task to post-task values.

6.2.2 Rhythm

Previous research has shown that L2 rhythm is influenced by L1 rhythm (White & Mattys, 2007). Dyads in this group were created using speakers who did not share an L1. As a result, the rhythmic properties of one language would be marked for speakers of the other language and, as per Goldinger's (1998) prediction, susceptible to convergence. This is further supported by analyses of pilot data of a separate set of speakers from the same language backgrounds which found convergence in rhythm after L1 – L2 interactions (Rao et al., 2011).

As noted in the hypotheses in Chapters 4 and 5, convergence would be characterized by either a decrease in the distance between the speaker pair's centroids or a reduction in the height of both the provider's and receiver's centroids. Conversely, divergence would be indicated by an increase in the distance of a speaker pair's centroids or an increase in the heights of both speakers' centroids. Evaluating convergence and divergence using EMS + centroid is described in Section 3.5.2.

Similar to the mixed dialect group, EMS + centroid would detect rhythmic differences specific to language and speaker pair. EMS + centroid may be capable of detecting rhythmic differences based on language differences (Rao & Smiljanic, 2011a). As noted in the introduction, Spanish shows less vowel and consonant reductions and deletions as compared to American English. Thus, AE may show more variable rhythm than Spanish as indicated by a higher centroid for the AE speakers than the Spanish speakers.

6.3 METHODOLOGY

Methods and stimuli are as described in the chapter on methodology, Chapter 3. Information specific to the mixed language group is provided below.

6.3.1 Participants

24 speakers (12 male) comprised the mixed language group. 12 of the 24 speakers were native Spanish (SP) speakers and 12 were native AE speakers. The Spanish speakers ages ranged from 19-23 years old and the AE speakers' ages ranged from 18-24 years old (mean = 21.14, SD = 1.64). Participants took an average of 33.16 minutes (SD = 15.26) to complete all four maps in the map task. This was 7.6 minutes longer than the mean of the native language and mixed dialect groups. Participants did not have any known speech or hearing impairments at the time of recording. They were either students (graduate or undergraduate) at the University of Texas at Austin or professionals living in the greater Austin area.

Spanish speakers were selected based on their speaking proficiency in English. Spanish speakers who had immigrated to the US before the age of 10 were excluded. Based on the background questionnaire (see Appendix C), SP speakers rated their English speaking proficiency at an average of 8.40 (SD = 1.56) on a Likert scale of 1-10 where 1 is least proficient and 10 is native-like proficiency. They also reported high proficiency in reading English on a similar scale (mean = 9.2, SD = 0.79). Besides English, two native SP speakers had experience with a second language via high school or college courses (French and Portuguese). Speakers were from five different Spanish-speaking countries: Cuba, Peru, Mexico, Colombia and Costa Rica.

Seven of the 12 AE speakers had spent all their lives in Texas; 5 others were from Connecticut, New York, Montana, Georgia, Louisiana and Alabama. One native AE speaker was Chinese but had moved to the US at the age of 2 and did not speak Chinese fluently. She also reported that Chinese was not spoken in her home. All but two had

some experience with a second language (Spanish, French, Swedish, American Sign Language and Chinese) via high school or college courses as part of a language requirement but were not fluent in any of these languages as indicated in the background language questionnaire.

Participants were assigned receiver or provider roles upon arrival to the lab. The first pair in this condition was assigned roles randomly and subsequent pairs were assigned the opposite roles to counterbalance the design. In this way, 3 Spanish speakers were providers (paired with AE receivers) and 3 were receivers (paired with AE providers).

6.4 RESULTS

The following section describes the results of the formant midpoint, IS and rhythm analyses in that order. For each section, the descriptive trends are discussed first followed by the statistical findings separated by sex. Alternative descriptive plots for vowels are also provided in Appendix E. Results from all statistical analyses are provided in Appendix D.

6.4.1 Midpoint formant analyses

6.4.1.1 Female speakers

Female F1 and F2 averages for each vowel separated by task and role are provided in Figure 6.1. It can be seen that receivers tended to maintain smaller values of F1 and F2 when compared to providers. F1 values for /i/ and /e/ are exceptions to this pattern. There are also some task-induced alterations. For example, the difference

between providers' and receivers' F2 values for /æ/ may have reduced from pre-task to post-task.

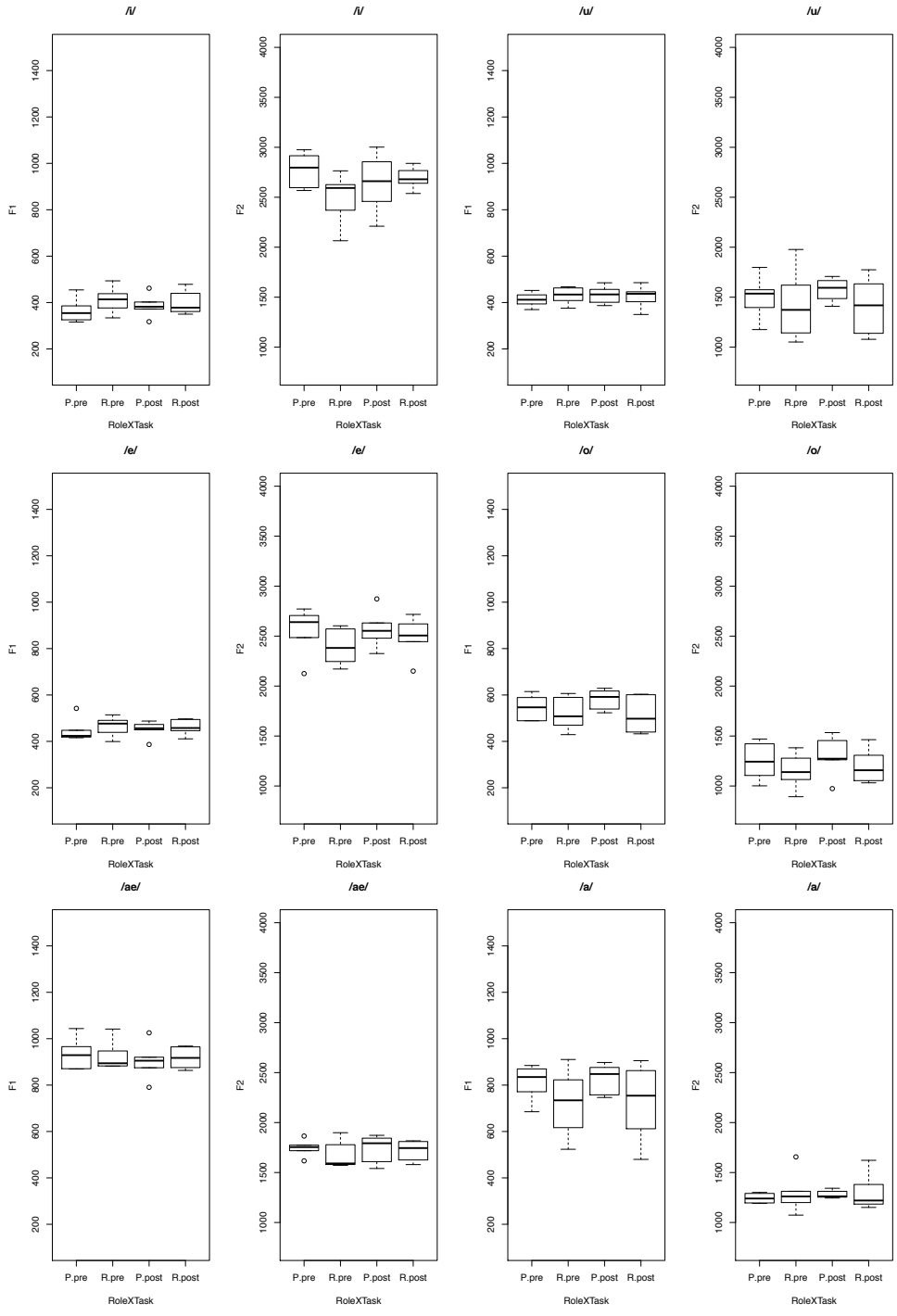


Figure 6.1: Average F1 and F2 values for all female vowels separated by role and task. Vowels are arranged in order of vowel height from top to bottom (/ɑ/ and /æ/ are listed as ‘a’ and ‘ae’ respectively in figure).

Mean female F1 and F2 values based on role and task for each language group are provided in Tables 6.1 and 6.2. All standard errors (SE) are listed in parentheses. The data suggest both AE and SP speakers altered their F1 and F2 values following the interaction. Providers, regardless of language background, increased their F1 values. Receivers, regardless of language background, increased their F2 values. Lastly, SP speakers maintained lower F1 and F2 values than AE speakers.

	Providers		**Receivers**	
Dialect	**Pre**	**Post**	**Pre**	**Post**
AE	605.747 (57.71)	614.136 (52.95)	605.458 (51.99)	592.304 (53.20)
SP	566.416 (47.93)	581.100 (46.30)	552.523 (42.75)	555.255 (43.66)

Table 6.1: Mean F1 values for female speakers separated by task, role and language.

	Providers		**Receivers**	
Dialect	**Pre**	**Post**	**Pre**	**Post**
AE	1855.865 (138.12)	1911.316 (139.21)	1779.512 (132.55)	1862.539 (148.09)
SP	1836.186 (165.21)	1789.579 (140.36)	1696.838 (139.75)	1738.735 (144.78)

Table 6.2: Mean F2 values for female speakers separated by task, role and language.

A 6 X 2 X 2 X 2 mixed-design ANOVA (evaluated with significance level alpha adjusted at 0.025), using F1 as a dependent variable (DV), vowel type (/æ, ɑ, i, u, e, o/) and task (pre or post) as within subjects factors and role (receiver or provider) and language (AE or SP) as between subjects factors revealed main effects of vowel ($F(5, 10) = 587.15$, $p < 0.001$) and language ($F(1, 72) = 4.288$, $p < 0.01$) and a three-way

interaction between vowel, language and role ($F(5,72) = 3.555$, $p < 0.01$). Post-hoc tests of simple effects using $\alpha = 0.004$ showed main effects of language and role for /ɑ/. Providers, regardless of L1, produced smaller /ɑ/ vowels than receivers. AE speakers produced lower /ɑ/ vowels than SP speakers (Figure 6.3).

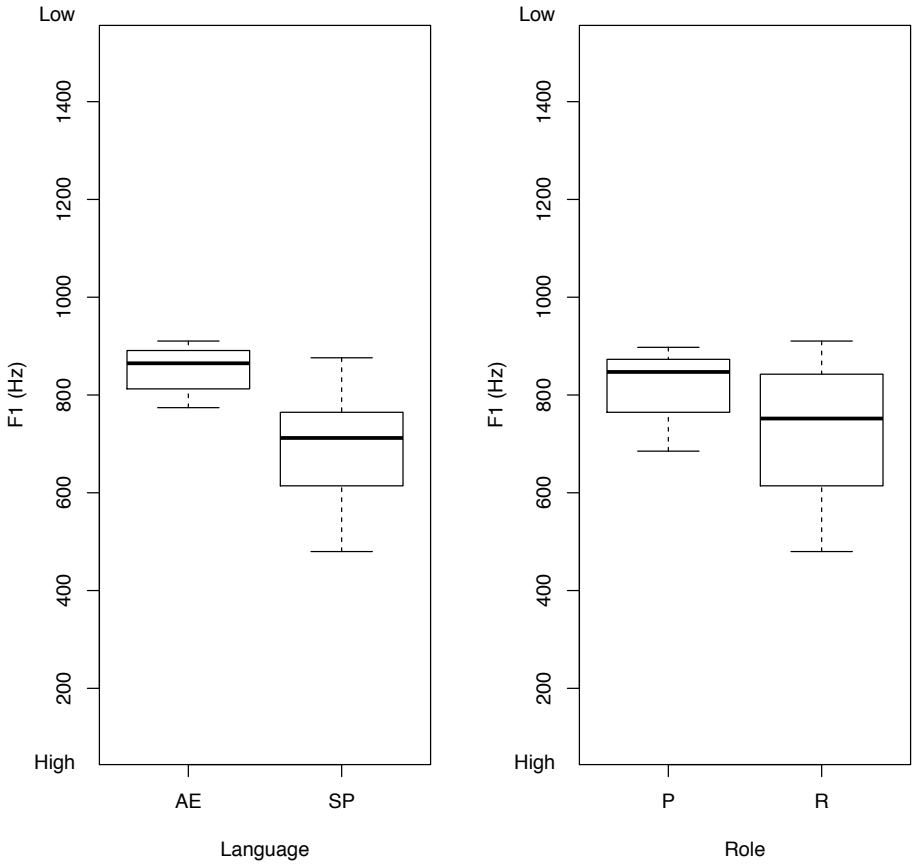


Figure 6.3: Female F1 values for /ɑ/ separated by language and role. 'Low' and 'high' indicate vowel quality.

Another mixed-design ANOVA with the same IVs but F2 as the DV revealed main effects of vowel ($F(5, 10) = 204.8$, $p < 0.001$), role ($F(1, 72) = 6.688$, $p < 0.025$) and language ($F(1, 72) = 8.141$, $p < 0.01$). There were no significant interactions. A plot of the main effect of role on F2 is provided in Figure 6.2, which shows that providers produced more fronted vowels than receivers.

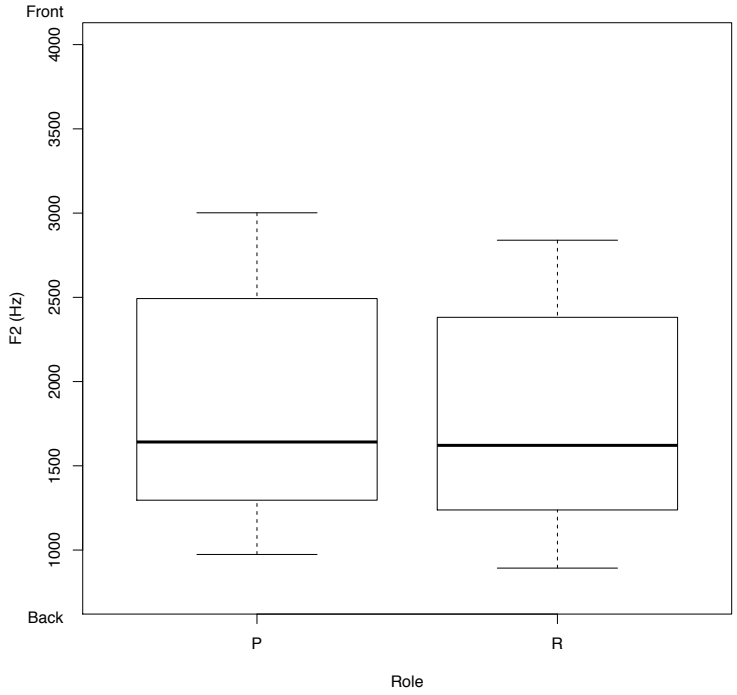


Figure 6.2: Female F2 values separated by role. 'Front' and 'back' indicate vowel quality.

A plot of the main effect of language on F1 and F2 is provided in Figure 6.3. It shows that AE speakers produced vowels with larger F1 and F2 values indicating lower and more fronted vowels than SP speakers.

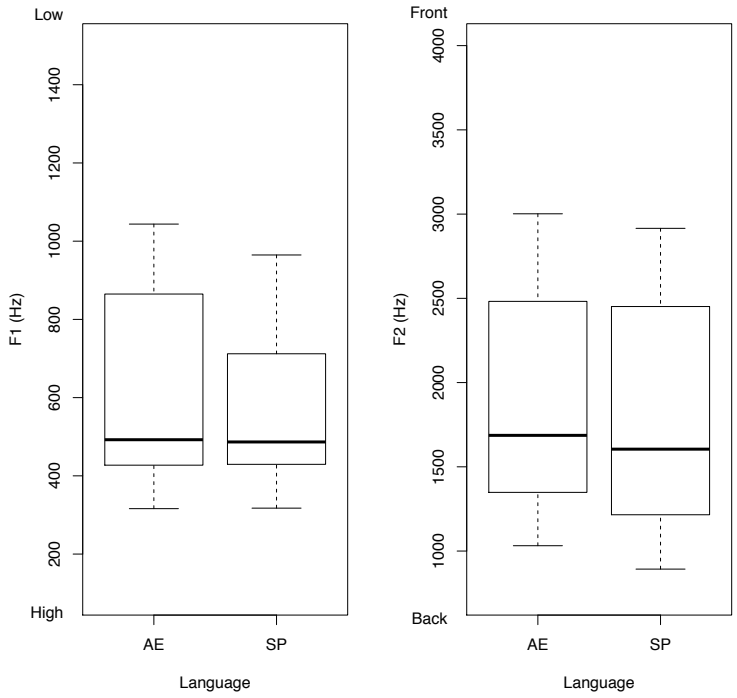


Figure 6.3: Female F1 and F2 values separated by language. 'Low', 'high', 'front' and 'back' indicate vowel quality.

6.4.1.2 Male speakers

Male F1 and F2 averages separated by vowel, task and role are provided in Figure 6.4. These do not show any consistent role-based or task-based differences in men's vowels.

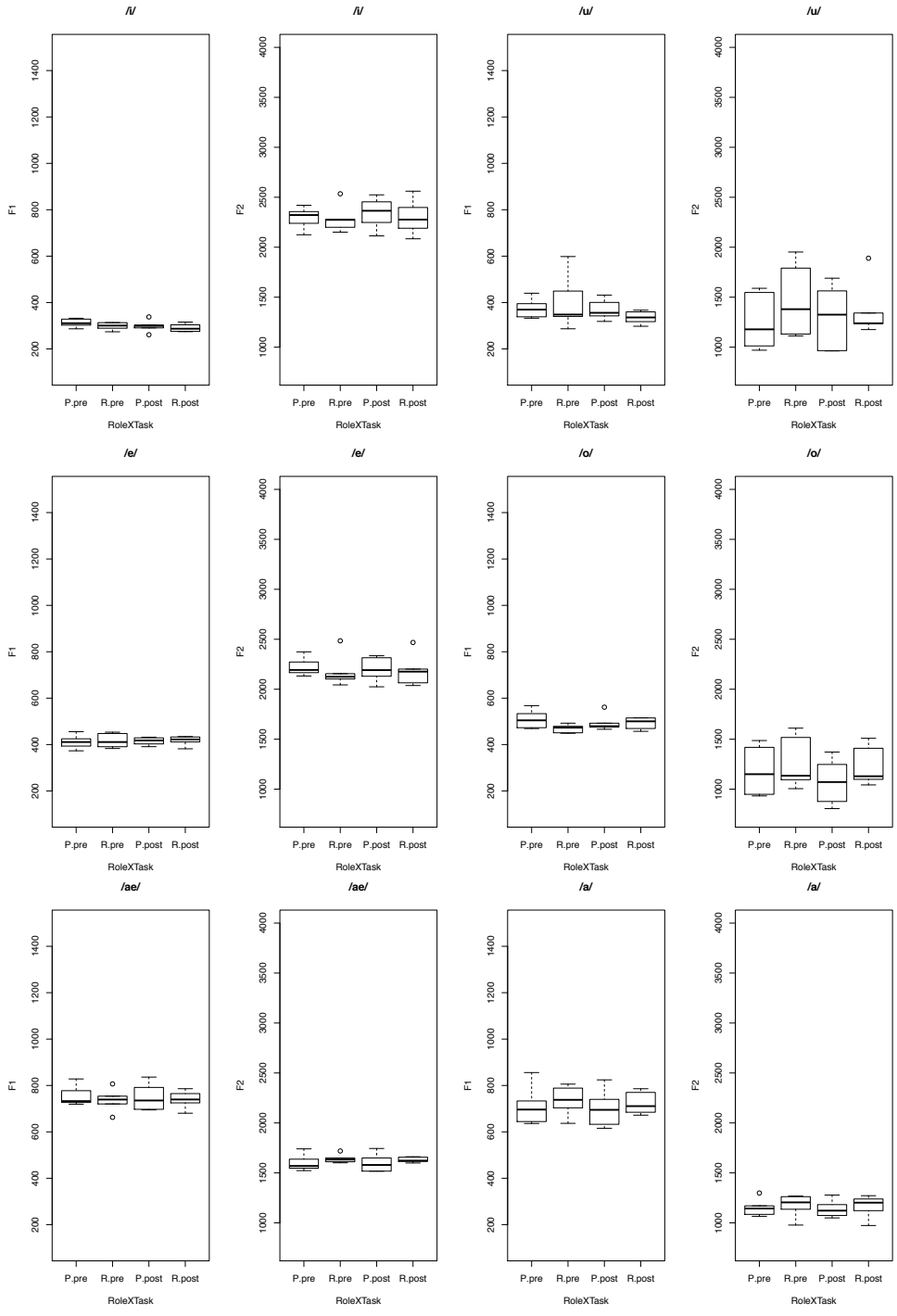


Figure 6.4: Average F1 and F2 values for all male vowels separated by role and task. Vowels are arranged in order of vowel height from top to bottom (/ɑ/ and /æ/ are listed as ‘a’ and ‘ae’ respectively in figure).

Mean male F1 and F2 values based on language, role and task are provided in Tables 6.3 and 6.4. All standard errors (SE) are listed in parentheses. Providers dropped their F1 values from pre-task to post-task regardless of language background. Language background affected how F2 values were altered from pre-task to post-task. For example, receivers who were AE speakers decreased their F2 values but receivers who were SP speakers increased their F2 values from pre-task to post-task.

	Providers		**Receivers**	
Dialect	**Pre**	**Post**	**Pre**	**Post**
AE	514.2494 (46.16)	508.5718 (44.19)	504.3989 (41.34)	505.4605 (44.89)
SP	508.7994 (36.54)	498.6961 (39.25)	512.7027 (44.35)	493.8751 (41.06)

Table 6.3: Mean F1 values for male speakers separated by task, role and language.

	Providers		**Receivers**	
Dialect	**Pre**	**Post**	**Pre**	**Post**
AE	1659.688 (114.99)	1693.188 (110.47)	1715.491 (100.83)	1669.876 (109.02)
SP	1571.936 (126.60)	1526.098 (139.75)	1611.945 (122.27)	1615.424 (121.09)

Table 6.4: Mean F2 values for male speakers separated by task, role and language.

A 6 X 2 X 2 X 2 mixed-design ANOVA, using male F1 values as a dependent variable (DV), vowel type (/æ, ɑ, i, u, e, o/) and task (pre or post) as within subjects factors and role (receiver or provider) and language (AE or SP) as between subjects factors revealed main effects of vowel ($F(5, 10) = 530.1$, $p < 0.001$) and task ($F(1, 72) = 47.07$, $p < 0.05$). Male speakers dropped their F1 values to create more raised vowels

after the interaction. No other main effects or interactions were noted. Mean F1 values pre- and post-task are provided in Figure 6.5.

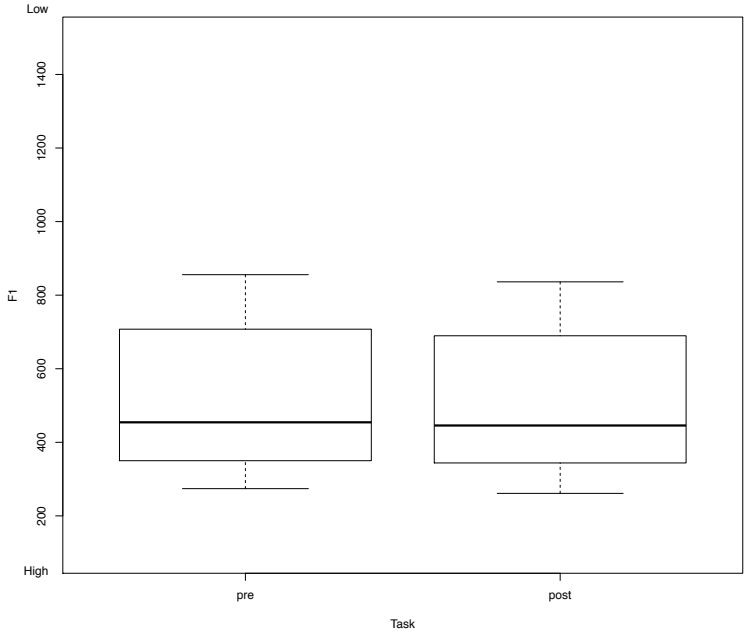


Figure 6.5: Male F1 values separated by task. 'High' and 'low' indicate vowel quality.

An analogous ANOVA with the same IVs as the F1 analysis but using male F2 values as the DV revealed main effects of vowel ($F(5, 10) = 199.7$, $p < 0.001$) and language ($F(1, 72) = 15.750$, $p < 0.001$) and a two-way interaction between vowel and language ($F(1, 72) = 7.940$, $p < 0.001$). Post-hoc tests of simple effects using $\alpha = 0.004$ showed that /u/ and /o/ were significantly different between the male SP and AE speakers. AE men's /u/ and /o/ were produced in a more fronted fashion than SP men, though generally, AE vowels were more fronted than SP vowels (Figure 6.6).

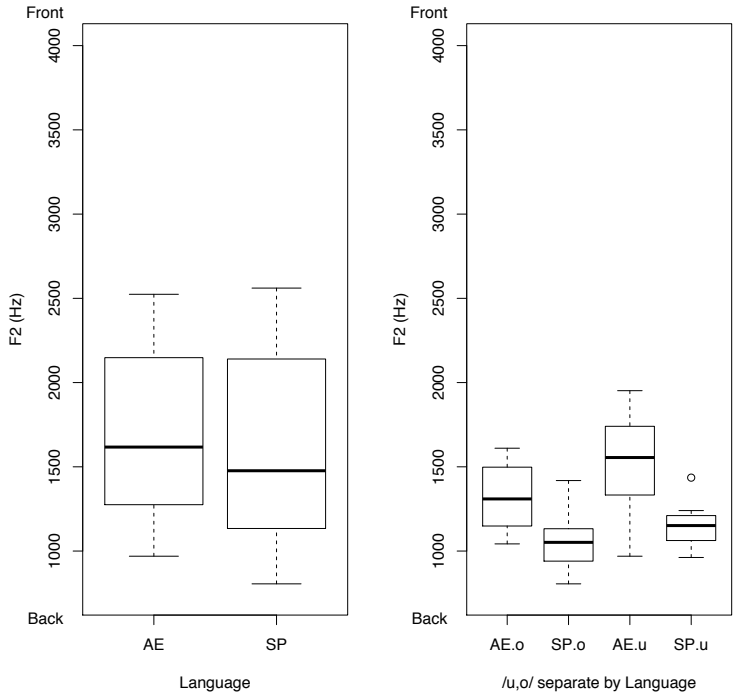


Figure 6.6: Male F2 means for vowels separated by dialect. 'Front' and 'back' indicate vowel quality.

6.4.2 Interlocutor similarity (IS)

Table 6.5 provides the means for all IS scores obtained via the bootstraps. It appears that all average scores, male sys-IS and spec-IS and female sys-IS and spec-IS, increased slightly from pre-task to post-task.

	Pre-task	Post-task
Female sys-IS	0.9912 (0.00)	0.9918 (0.00)
Female spec-IS	0.9967 (0.00)	0.9970 (0.00)
Male sys-IS	0.9848 (0.00)	0.9880 (0.00)
Male spec-IS	0.9969 (0.00)	0.9975 (0.00)

Table 6.5: means of significant IS scores for male and female speakers (SE in parentheses are 0 up to two significant digits).

Bootstrap analyses of two repeated-measures ANOVA with sys-IS as DV and task (pre-task, post-task) as IV were run separately on the male and female data. Sys-IS and spec-IS are a measure of similarity between the receiver and provider. As a result, role and language background cannot be tracked anymore. These bootstraps revealed significant effect of task for male speakers. 97.5% CIs are provided in Figure 6.7. Female speakers did not show a main effect of task.

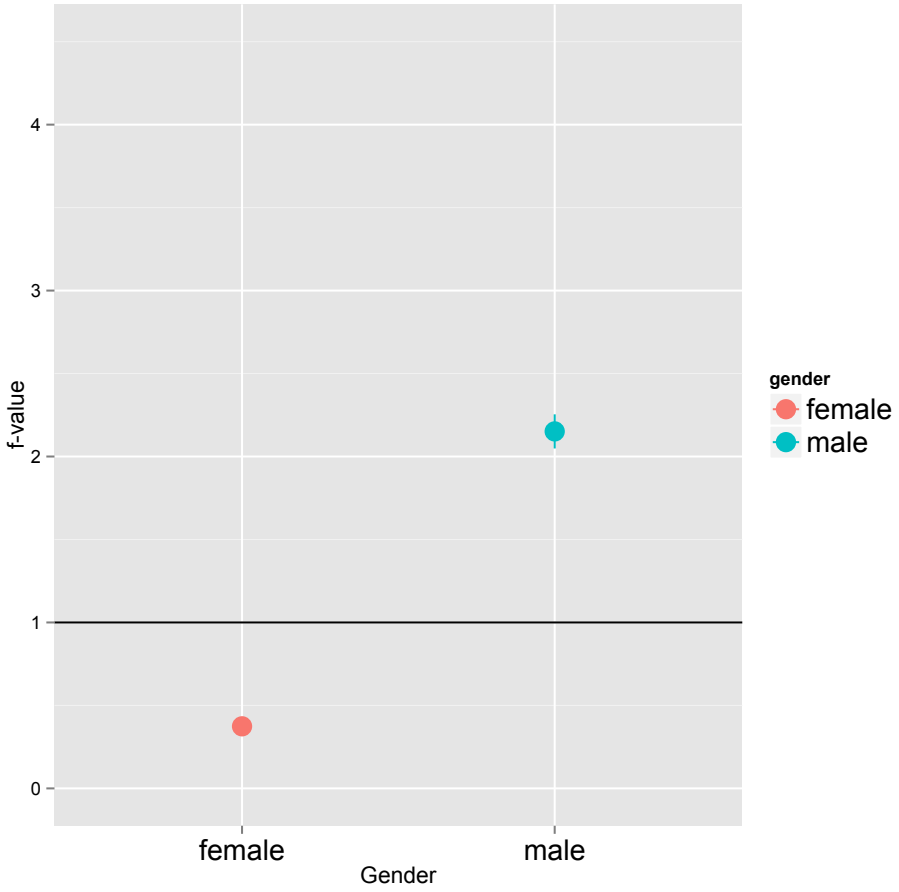


Figure 6.7: 97.5% CI for sys-IS F-values for male and female speakers. Red points indicate female f-values and blue indicate male f-values.

Another pair of bootstrap analyses of repeated-measures ANOVAs using spec-IS as DV and task and vowel as IVs revealed main effects of task for both male and female speakers (Figure 6.8). A significant interaction between vowel and task for the male speakers was also found. However, these effects, particularly those for women were small. Women also showed a very small increase (a change of 0.0003) in vowel-specific similarity via spec-IS after the map task (Table 6.5). This small of an increase combined with the mean F-value of the spec-IS analysis being close to 1 suggests that this effect

was not meaningful and therefore will not be discussed. Vowel specific differences of spec-IS are provided in Figure 6.9 where it can be seen that /æ, u, e, o/ increased in similarity, /ɑ/ decreased in similarity and /i/ did not show any changes in spec-IS scores signaling convergence, divergence and maintenance respectively.

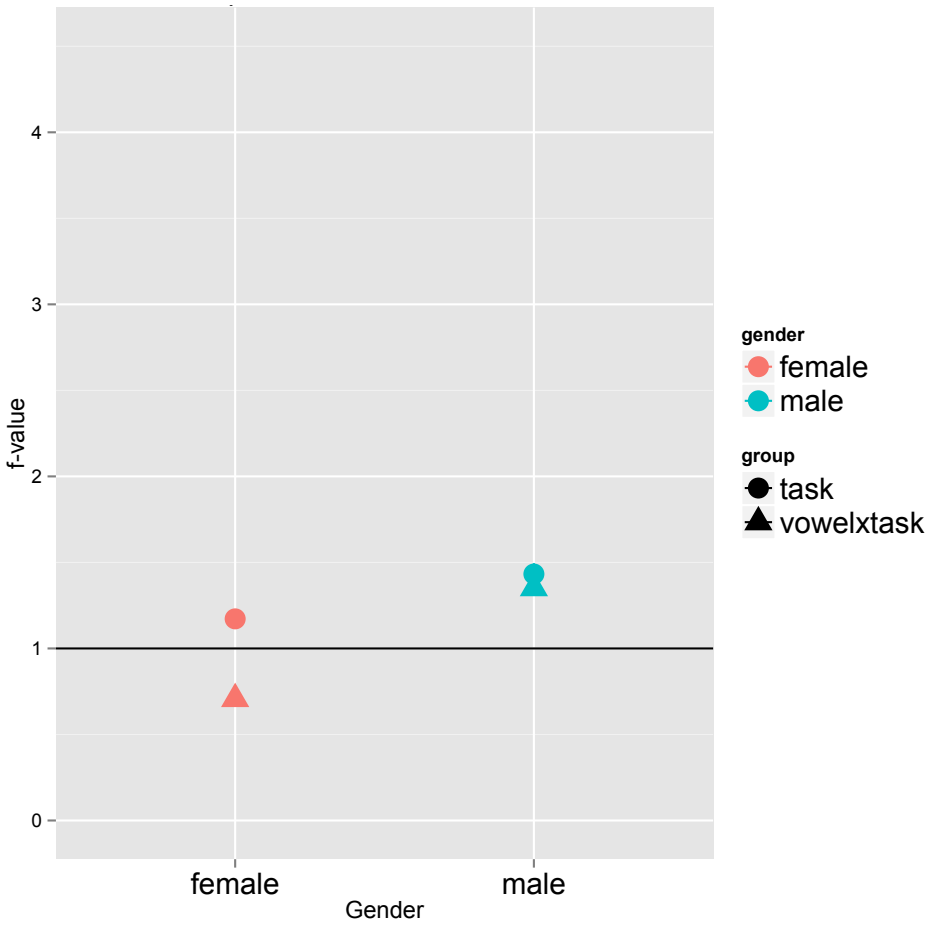


Figure 6.8: 97.5% CI for spec-IS F-values for male and female speakers. Red points indicate female and blue indicate male f-values. Circles indicate task f-values whereas triangles indicate f-values for the vowel X task interaction.

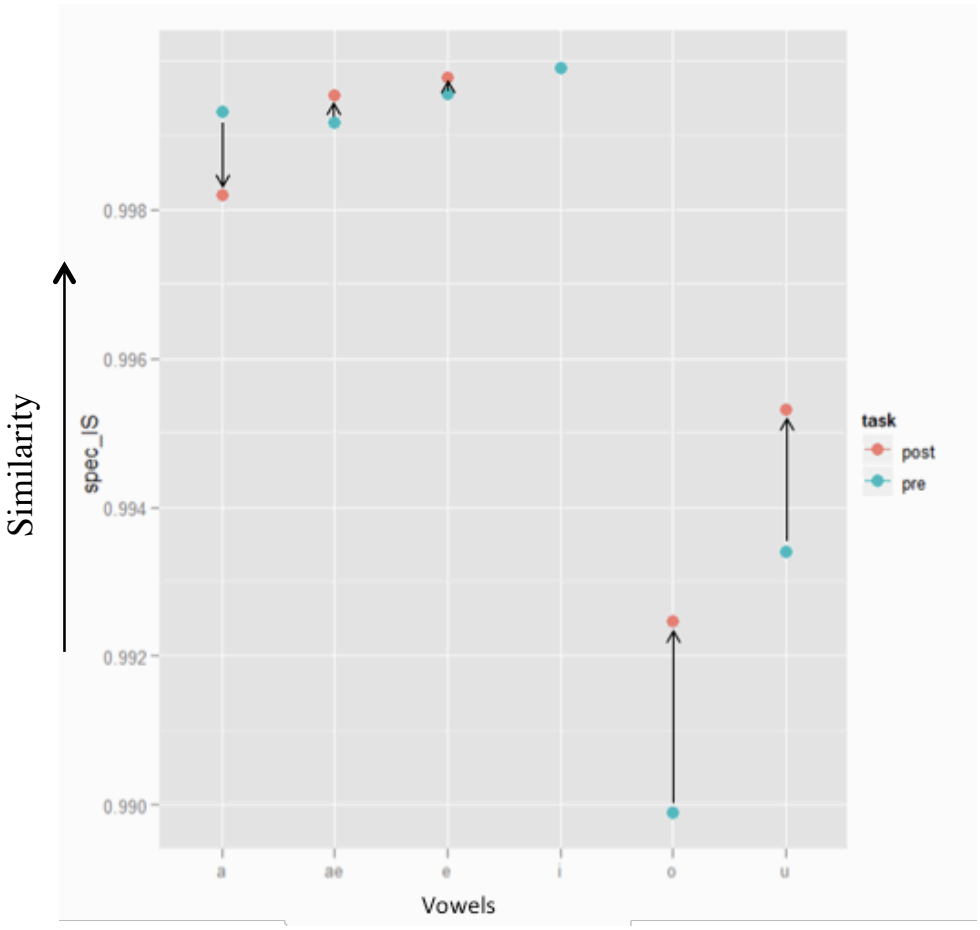


Figure 6.9: means of bootstrapped spec-IS values by vowel for male speakers. /ɑ/ and /æ/ are listed as 'a' and 'ae' respectively in this plot. Red points indicate post-task means and blue points indicate pre-task means.

6.4.3 Rhythm

6.4.3.1 Female speakers

Female centroid averages are provided in Table 6.6 (SE are in parenthesis). These averages show that SP speakers decreased their centroids from pre-task to post-task. AE

speakers increased their centroids from pre-task to post-task. For SP speakers, receivers had a lower centroid than SP providers.

	Providers		**Receivers**	
	Pre-task	**Post-task**	**Pre-task**	**Post-task**
AE	8.835 (0.53)	8.584 (0.26)	8.580 (0.35)	8.682 (0.31)
SP	9.290 (0.28)	8.792 (0.26)	8.461 (0.33)	8.412 (0.30)

Table 6.6: Mean rhythm centroids for female speakers separated by task, role and language.

A 2 X 3 X 2 X 2 nested factorial ANOVA using role (receiver or provider), speaker pair (speaker pairs 1-3) and task (pre or post) and language (SP or AE) as IVs and EMS + centroid values as the DV was run on the female data. As with previous analyses, significance levels were adjusted to $\alpha = 0.025$. Because native language as a factor was also included in these analyses, speaker pairs had to be coded in a manner that collapsed two speaker pairs into one. Thus, each speaker pair contained one $Provider_{Spanish}$-$Reciever_{AE}$ and one $Provider_{AE}$-$Reciever_{Spanish}$ combination.

The analysis of female centroids showed an interaction between role x speaker pair x language ($F(2, 120) = 10.1730$, $p < 0.001$). No main effects were noted for this analysis. Post-hoc tests of simple effects using $\alpha = 0.008$ showed that AE speakers from female speaker pair 3 had significantly different centroids based on role. The female AE receivers had lower centroids (mean = 9.456; SE = 0.29) than female SP receivers (mean = 7.683; SE = 0.392).

6.4.3.2 Male speakers

Male centroid averages are provided in Table 6.7 (SE are in parenthesis). These show that male receivers regardless of native language raised their centroids from pre-task to post-task.

	Providers		**Receivers**	
	Pre-task	**Post-task**	**Pre-task**	**Post-task**
AE	8.294 (0.28)	8.224 (0.31)	8.461 (0.29)	9.194 (0.27)
SP	8.169 (0.33)	8.101 (0.38)	7.795 (0.35)	7.847 (0.30)

Table 6.7: Mean rhythm centroids for male speakers separated by task, role and language.

Another 2 X 3 X 2 X 2 nested factorial ANOVA of the male data with the same IVs and DV showed main effects of speaker pair ($F(2,40) = 5.588$, $p < 0.01$) and language ($F(1,120) = 12.480$, $p < 0.001$). SP centroids were lower 7.978 (SE = 0.17) than AE centroids 8.543 (SE = 0.14). One interaction, role x speaker pair x language ($F(2,120) = 13.363$, $p < 0.001$), was also noted. Post-hoc tests of simple effects using $\alpha = 0.002$ showed that for male speaker pair 1, centroids were significantly different for SP and AE providers and receivers. These means are provided in Table 6.8 (SE in parentheses) and show that AE providers had lower centroids than SP providers and AE receivers had higher centroids than SP receivers.

	Providers	**Receivers**
AE	8.69(0.33)	9.21(0.31)
SP	9.79(0.31)	7.13 (0.31)

Table 6.8: Average centroid values for male speakers from speaker pairs 1 separated by role.

Post-hoc Bonferroni tests also showed that speaker pairs 1 were significantly different from 2 and 3. Their centroids are provided below in Table 6.9 (SE in parentheses).

	Centroids
Speaker pairs 1	8.708 (0.19)
Speaker pairs 2	8.103 (0.21)
Speaker pairs 3	7.972 (0.18)

Table 6.9: Mean speaker pairs rhythm centroids.

6.4 DISCUSSION

6.4.1 Vowels

Formant analysis for female data revealed differences due to role and language. The /ɑ/ vowel was lower for female AE speakers than for female Spanish speakers. Furthermore, all female AE speakers' vowels demonstrated higher F1 and F2 values than Spanish speakers' suggesting that the AE women produced vowels that were lowered and fronted more than the Spanish women produced. Female providers generally fronted their vowels more than female receivers fronted their vowels. Furthermore, these providers also produced lower /ɑ/ vowels than female receivers produced.

An effect of interaction on vowels due to the map task was found only for male speakers who produced lowered vowels (higher F1 values) after the task than before it. However, interpreting this lowering is difficult because it occurred regardless of speaker role. The IS measures clarify this adaptation by noting that men altered their vowels after the interaction. The effect of task on sys-IS was larger compared to spec-IS, suggesting that convergence was better detected when changes across the entire vowel space were

considered. Spec-IS scores support this indication of convergence by showing that for men all vowels except for /ɑ/ and /i/ converged.

The Spanish vowel system consists of /a, e, i, o, u/; /æ/ and /ɑ/ are absent from the SP vowel inventory and of particular interest in this study. Based on spec-IS, both of these vowels showed task-induced adaptations in male productions: /æ/ converged while /ɑ/ diverged. Since neither of the vowels is in the Spanish vowel inventory, they were expected to show consistent adaptation patterns, *i.e.* either divergence or convergence. The reasons for differing patterns of convergence for these vowels are unclear. Figures 6.10 and 6.11 suggest that speakers from both linguistic backgrounds altered their vowels to create divergence in /ɑ/ and convergence in /æ/. Both AE and SP male speakers altered their F1 and F2 values for these vowels suggesting that speakers from both linguistic backgrounds noticed the difference in the vowels produced by their partners.

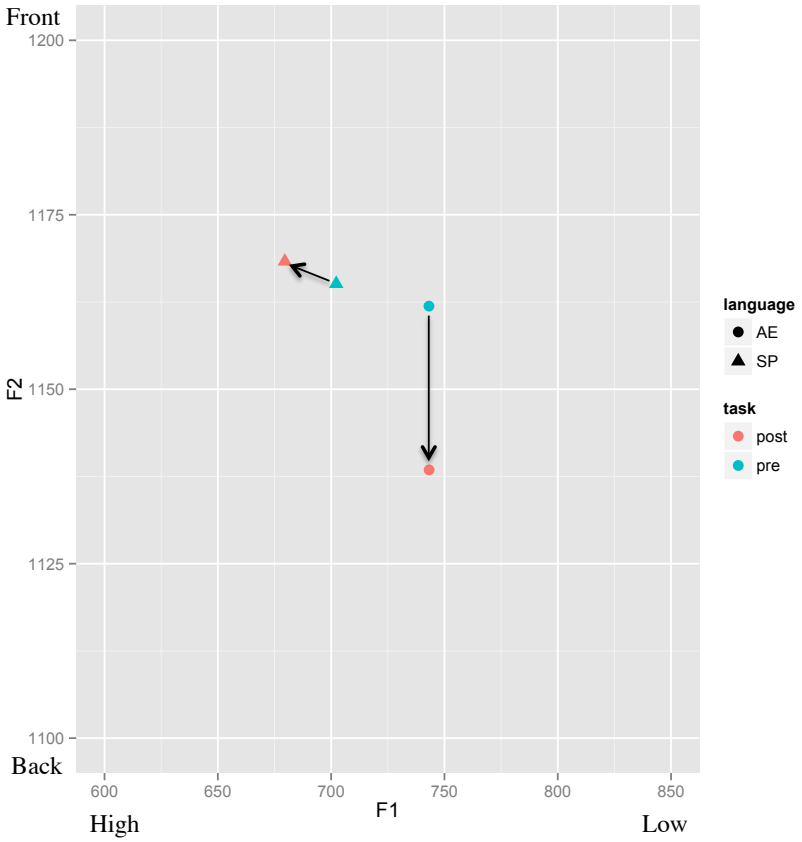


Figure 6.10: Mean F1 and F2 for /ɑ/ separated by task and language (male speakers).

For /ɑ/, AE speakers were responsible for most of the divergence; they decreased the F2 values of their /ɑ/ vowels to created more backed versions. Spanish speakers increased F2 and decreased F1 values of their /ɑ/ vowels to create more fronted and raised versions of the same vowel. This alteration resulted in divergence.

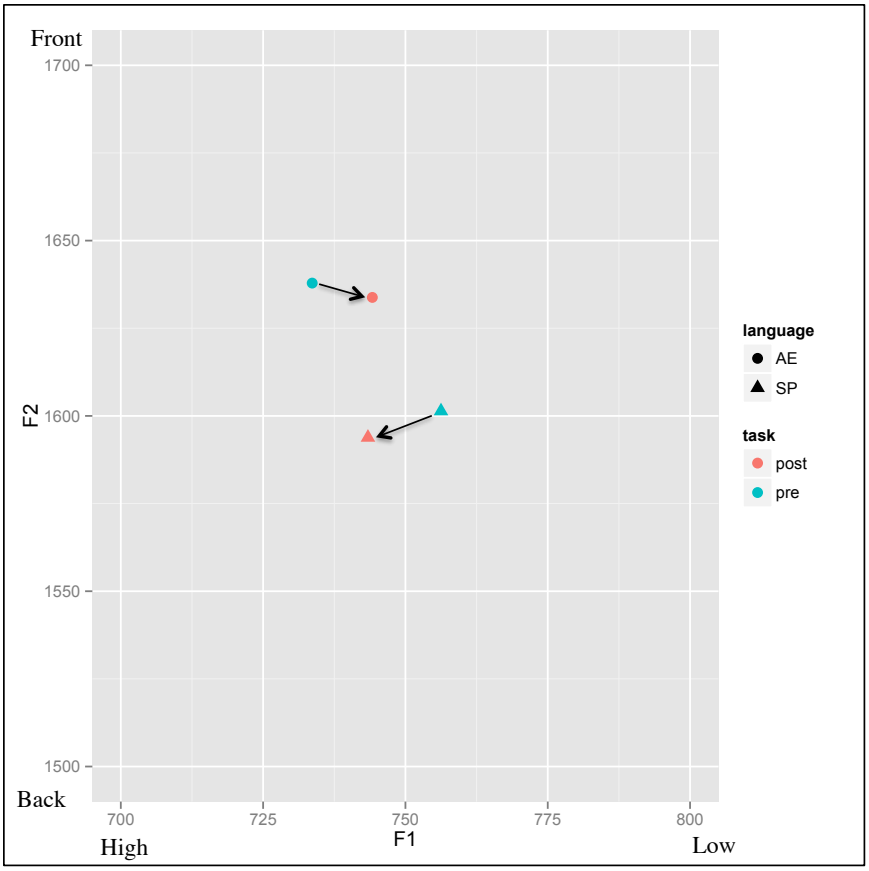


Figure 6.11: Mean F1 and F2 for /æ/ separated by task and language (male speakers).

Similar to the adaptations noted in /ɑ/, men from both linguistic backgrounds altered their /æ/ vowels resulting in convergence. Spanish speakers raised their version whereas AE speakers lowered theirs. The convergence patterns gave rise to a more fronted /æ/ for AE men and a more backed version for Spanish men. Hence, the prediction that marked vowels would exhibit convergence was both confirmed and contradicted by the adaptations noted in /æ/ and /ɑ/. Additional research is necessary to identify the reasons for the divergence noted in /ɑ/.

6.4.2 Rhythm

Another a-priori prediction had been that rhythm convergence would be noted in male and female dyads. This prediction was not confirmed because no effect of task was noted. Reasons to not adapt to a conversational partner who does not share L1 would be different for L1 AE and L1 Spanish speakers. Changing the L1 AE rhythm to be more like Spanish would entail less vowel and consonant variation, which could possibly lead to a non-native AE sounding rhythm, thus deterring AE speakers from adapting. On the other hand, given that L1 rhythm affects L2 rhythm (White & Mattys, 2007), Spanish speakers may have been incapable of altering their rhythmic properties any further due to L1 (Spanish) interference of L2 (AE).

The results revealed individual differences in male rhythm patterns. Role-based differences in rhythm were also noted for both male and female dyads but these differences are difficult to interpret without task involved as a factor. Speaker pairs showed rhythmic differences based on the role they were assigned that they maintained regardless of the verbal interaction.

Spanish and AE are considered prototypical syllable- and stress-timed languages, respectively. However, the findings from the current study did not detect a straightforward distinction between Spanish vs. English rhythm. EMS + centroid noted less variable rhythm (lower centroid) for men with Spanish L1 than for men with AE L1. This agrees with the previous findings that EMS + centroid detects increased rhythmic variability (or a higher centroid) due to consonant and vowel deletions and reductions that are more likely in English than Spanish (Rao & Smiljanic, 2011b). However, similar effects were not noted for female data. Listening to male and female data leaves the impression that women were more accented than men. It is possible that in women, AE speakers were switching to clear speech to aid intelligibility. This would have led to a

lowered centroid that was comparable with the SP centroids. Rhythm distinctions are discussed further within the context of the larger study in the general discussion.

6.4.3 General trends

Generally, women who were far apart in terms of linguistic distance did not show adaptations in vowels or rhythm after the interaction. Men in the same group showed vowel-specific adaptations but did not show any rhythm adaptations. All speakers (native and non-native) were proficient readers and writers of English. However, the non-native speakers reported mid-high proficiency in speaking English. Separating the speaking proficiency based on sex revealed that non-native men self-rated as more proficient (9/10) than non-native women (7.5/10). Perhaps, greater speaking proficiency allowed men to adapt their vowels. Lewandowski (2012) found that speakers who were more adept at native-like pronunciation were more likely to converge in a temporal prosodic measure than those who were not as adept. The current finding suggests that speakers who are adept at speaking an L2 are able to adapt at the segmental level too.

Chapter 7: Rhythm Convergence During Interaction

7.1 Introduction

While most of the research on PC has focused on adaptations taking place to interlocutors' speech patterns by comparing controlled speech tokens before and after an interaction, few studies have examined adaptations during the interaction task itself (Krivokapic, 2013; Lewandowski, 2012; Pardo, Jay, & Krauss, 2010). For example, Pardo, Jay, & Krauss (2010) examined changes in articulation rates during a map task. In this study, either the provider or the receiver from each dyad was specifically instructed to imitate their partner's speech. Time-series cross-correlation analyses, which compared similarity in speakers' articulation rates, found no consistent changes; however listener judgments indicated convergence. Specifically, they judged utterances to be more similar between interlocutors when female receivers and male providers were instructed to imitate their partners. Similarly, Krivokapic (2013) examined the rhythm of IE-AE speaking pairs during a synchronized reading task in which she used tokens from the beginning and end (or early and late) during the task and found convergence in the rhythm on one out of four dyads. Finally, Lewandowski (2012) examined PC in native and non-native speakers using read and conversational speech. She used two separate comparisons: early and late words from a matching picture (diapix) task and the same target words produced pre-task and post-task as part of a reading list. Although she reported convergence in early and late dialogue comparisons, she also found that convergence patterns in citation speech did not predict similar patterns in conversational speech from the task itself. The convergence noted in the early and late part of the dialogue was not noted in the pre-task and post-task tokens.

This chapter explores the emergence of PC in rhythm as tracked across the time course of the task. In this study, pre- and post-task data is citation style speech and the

map task data is conversational. Speaking style changes across these two conditions may create confounds in the analyses. Lewandowski's (2012) results suggest that PC was demonstrated in controlled pre-task/post-task analyses and in early and late tokens from the during-task dialogue. However, the results from these two analyses were not correlated suggesting that pre-task/post-task adaptations did not imply during-task adaptations. Furthermore, recording stimuli and speaking style (citation vs. conversational) can be a large source of rhythm variation (Wiget et al., 2010; Tilsen & Johnson, 2008). Pre-task and post-task items were the same set of read sentences but the map task items were free-form conversations that were not controlled. To ensure that conversational style and variable stimuli from the map task did not contribute to differences in PC, the pre-task and post-task data were first analyzed separately. These analyses did detect task-induced adaptations to speaker's vowels and rhythm: native AE speakers who were men converged in rhythm, speakers who spoke differing dialects of English showed both convergence and divergence and speakers who did not share the same L1 showed no adaptations. The map data is now included along with pre- and post-task tokens to examine rhythm adaptations as they develop over the course of the interaction with the pre-task tokens serving as baseline and post-task tokens serving as the end result. By analyzing speech during the map task interaction, the pattern of rhythm changes noted in the pre-task/post-task data may be elucidated further. As speakers did not reliably reproduce all the target vowels during the map task, the emergence of vowel convergence could not be examined in a similar manner. This happened often in cases where a landmark item was not included on a partner's map forcing interlocutors to use landmarks that may not have contained the target vowels. Analysis of changes to rhythm included tokens from pre-task and post-task recordings as well as sentences from the conversation during the map task itself. The during-task tokens from the map task are

included now that adaptation patterns from the more controlled pre- vs. post-task data are known. Considerations for segmentation of the map task conversation are outlined in the methodology section (7.3) below.

7.2 HYPOTHESES

7.2.1 Rhythm

Including data from the task itself should help clarify the trajectory of change to the rhythmic properties of the interlocutors' speech. Based on the findings from the analysis of the pre- and post-task data, it was predicted that the male dyads would show increasing adaptations in rhythm over the time course of the map task. Moreover, based on the rhythm results of Experiments 1, 2 and 3, language background was expected to interact with rhythm PC for male speakers during the interaction. The native language group would converge, the mixed dialect group would diverge and the mixed language group would show little or no change in prosodic rhythm across map interactions. These adaptations would be noted by EMS + centroid movements as outlined in Section 3.5.2. Female dyads would not show any notable changes in their rhythmic pattern regardless of language background.

7.3 METHODOLOGY

In order to compare changes taking place over the course of the interaction, speaker pairs who did not complete all four maps were excluded from the analysis. This resulted in two out of six male speaker pairs being dropped from the native language and mixed dialect language conditions who were stopped due to time considerations after completing 3 out of 4 maps. All male speaker pairs in the mixed language conditions and

all female speaker pairs were retained. Male speakers spent an average of 28.01 minutes (SD = 14.38) and female speakers took an average of 22.01 minutes (SD = 12.84) completing all four maps. Task was coded as 6 events; pre-task, map 1, map 2, map 3, map 4 and post-task.

Segmentation of map task utterances was done based on turn-taking (Edlund, Heldner & Gustafson, 2005). An utterance was marked once the speaker paused. In some instances, there was acknowledgement from the partner such as 'yes' or 'uh-huh.' In other cases, the partner was silent or repeated the utterance. Utterances where a speaker was interrupted by his or her partner and those that were less than 2 seconds were omitted from the analyses (personal communications with Sam Tilsen and Andrew Lotto). Furthermore, sentences that included noise from brushes against the microphone or paper rustles were also omitted.

Sentences from the map task differed from the pre-task and post-task tokens in some notable ways. Firstly, these sentences were either in the form of statements or questions. Secondly, incomplete statements such as false starts or hesitations and statements with ellipses or filled pauses were also included. Finally, not all sentences had landmark items in them. In contrast, the pre-task and post-task tokens were always complete statements, which contained landmark items. The EMS + centroid was calculated for these sentences using the methodology outlined in Chapter 3.

7.4 RESULTS

As with the previous chapter, because the male and female analyses were conducted separately, the alpha-level was adjusted using Šidák correction to 0.025. Results from all statistical analyses are provided in Appendix D.

7.4.1 Female speakers

Three hierarchical linear models (HLM) with random intercept were used to analyze rhythm changes within dyads of each language group. HLMs were used in order to track changes in rhythm over time. In each case, the *Akaike information criterion* (AIC) was used to evaluate the best model for each dataset. For each model, the centroid was used as the DV and task (pre, map1, map2, map3, map4 or post), role (receiver or provider) and speaker (speaker pairs 1-6) were used as IVs. For models of the mixed dialect and mixed language groups, language (AE or IE or SP) was also included as a factor.

The plot in Figure 7.1 below shows the centroid means of female speakers separated by role and task in each language group separately. This plot does not reveal a noticeable trend of any rhythm adaptations in female centroids across maps. In the native language group, providers have slightly lower centroids than receivers. This trend appears to be reversed in the mixed language group.

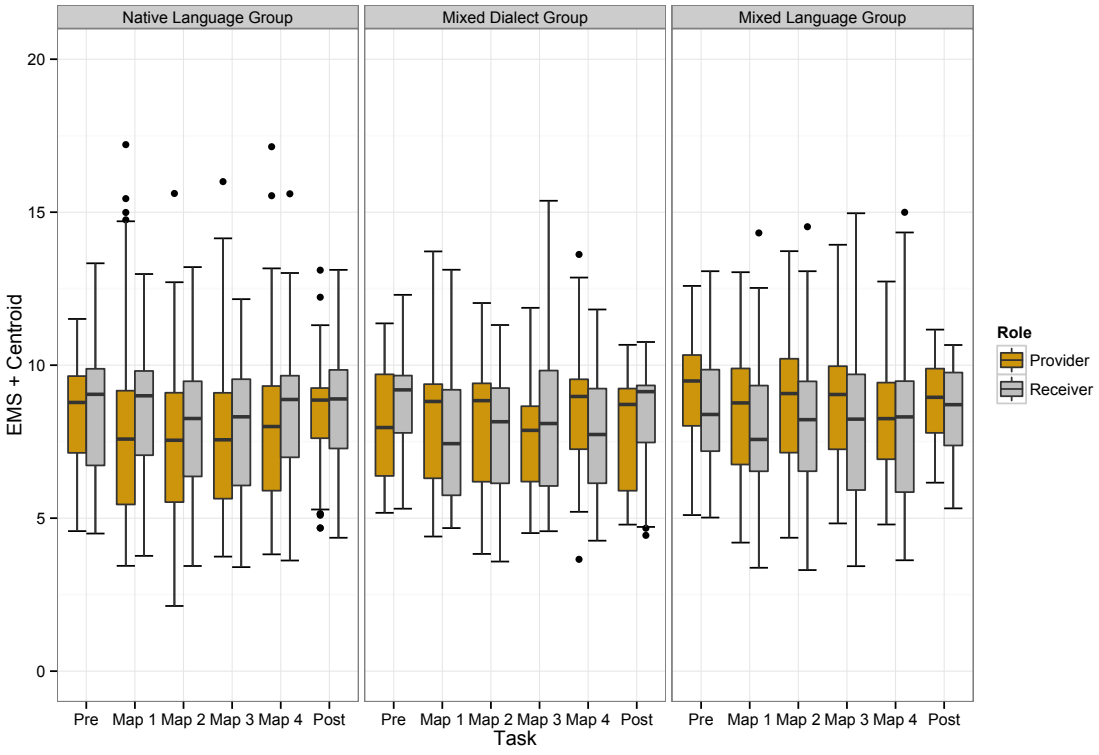


Figure 7.1: Mean rhythm centroids for female speakers separated by task and language.

The means for all female speaker pairs from the native language group are provided below in Table 7.1 (all standard errors (SE) are provided in parentheses). Analysis of the female data revealed a main effect of role in the native language group (significant coefficients, t-values and p-values are provided below in Table 7.2. Post-hoc Bonferroni tests showed that speaker pairs 3, 4, 5 and 6 were significantly different from speaker pair 1 and speaker pairs 5 and 6 were significantly different from 2. No other main effects or interactions were noted.

	Average centroid values
Speaker pair 1	7.337 (0.11)
Speaker pair 2	7.412 (0.14)
Speaker pair 3	7.924 (0.13)
Speaker pair 4	7.892 (0.13)
Speaker pair 5	8.169 (0.21)
Speaker pair 6	8.376 (0.17)

Table 7.1: Means of all speaker pair specific female rhythm centroids from the native language group.

	β-value	SE	t-value	p-value
Intercept	7.187	0.11	67.217	0.00
Speaker pair	0.23	0.09	2.41	0.01

Table 7.2: Significant coefficients, t-values and p-values for the female native language model

The analysis of the mixed dialect group also revealed a main effect of role (significant coefficients, t-values and p-values are provided below in Table 7.3). Providers (mean = 7.903; SE = 0.08) generally maintained a lower centroid than receivers (mean = 8.125; SE = 0.09).

	β-value	SE	t-value	p-value
Intercept	7.861	0.10	80.762	0.00
Role	1.60	0.54	2.93	0.003

Table 7.3: Significant coefficients, t-values and p-values for the female mixed dialect model

Lastly, the mixed language group did not reveal any significant effects or interactions for female speakers. No other main effects or interactions were noted in the native language, mixed dialect or mixed language groups.

7.4.2 Male speakers

The plot in Figure 7.2 below shows rhythm centroid means of male speakers separated by role and task for each language group separately. No clear trends in rhythm adaptation across maps are seen in these plots. However, providers show slightly lower centroids compared to receivers in all three language conditions.

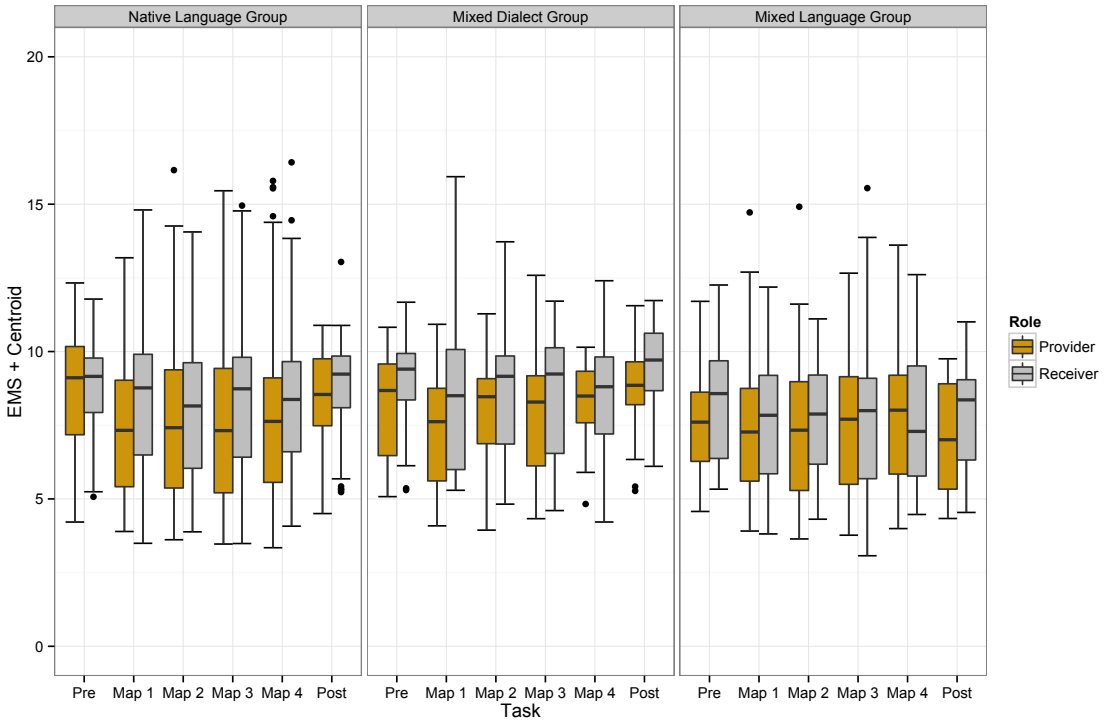


Figure 7.2: Mean rhythm centroids for male speakers separated by task.

The centroid means for all male speaker pairs are provided below in Table 7.4 (all SEs are provided in parentheses). Three HLMs that used the same DV and IVs were run on the male data separated by language condition. A significant main effect of speaker pair for the native language group was noted (significant coefficients, t-values and p-values are provided below in Table 7.5). Post-hoc Bonferroni tests showed that speaker pairs 1 and 2 were significantly different from speaker pairs 3 and 4 (speaker pairs 5 and 6 were dropped due to incomplete map tasks).

	Average centroid values
Speaker pair 1	7.660 (0.13)
Speaker pair 2	7.723(0.14)
Speaker pair 3	8.628 (0.15)
Speaker pair 4	8.511 (0.14)

Table 7.4: Means of all speaker specific male rhythm centroids from the native language group.

	β-value	**SE**	**t-value**	**p-value**
Intercept	7.622	0.12	62.719	0.00
Speaker pair	0.25	0.09	2.60	0.01

Table 7.5: Significant coefficients, t-values and p-values for the male native language model.

The analysis of male pairs from the mixed language group revealed a main effect of speaker (significant coefficients, t-values and p-values are provided below in Table 7.7). All speaker means are provided in table 7.6.

	Average centroid values
Speaker pair 1	8.012 (0.08)
Speaker pair 2	7.794 (0.08)
Speaker pair 3	7.534 (0.08)

Table 7.6: Means of all speaker specific male rhythm centroids from the mixed language group.

	β-value	SE	t-value	p-value
Intercept	8.146	0.30	26.281	0.00
Speaker	-0.648	0.25	-2.535	0.011

Table 7.7: Significant coefficients, t-values and p-values for the female mixed dialect model.

No other main effects or interactions were noted in the three language conditions.

7.5 DISCUSSION

Analysis of the entire dataset comprised of pre-, map and post-task tokens showed speaker pair-specific differences in the rhythm of women in the native language group and the mixed dialect group. Furthermore, providers from the native language group maintained lower centroids than receivers, thereby indicating a less variable rhythm than the receivers. It is likely that providers demonstrated less variability in their rhythmic patterns because they were typically stating the instructions. On the other hand, receivers asked questions and made statements when summarizing the routes. This would result in more amplitude fluctuations leading to a variable rhythmic pattern. Because there is not an interaction with task, this more stable rhythm in providers as compared to the receivers probably did not arise during the interaction. Thus, for women in the native

language group, speakers maintained a rhythmic difference due to the role they were assigned but did not show rhythmic adjustments due to the interaction.

Surprisingly, effects of role or task was not detected in male speakers' data either. Given the previous chapters' results, the prediction had been that men (but not women) would show changes in their rhythmic patterns across maps. However, including the interaction data did not show rhythmic adaptations across dialogue partners. It is important to note that the pre-task/post-task data was comprised of controlled sentences, whereas the map task consisted of freeform conversations that included more than just statements. Perhaps the amplitude distributions of the sentences from the map task fluctuated so greatly that any effects noted in the pre-task/post-task data were obscured.

Another explanation as to why no rhythm adaptations were noted when map data was included points to the large variation in time spent completing the map task. Even though the number of maps was held constant for everyone, some dyads finished the task in as little as 10 minutes while others took as much as 57 minutes to finish. This may have led to certain dyads adapting more than others resulting in an overall lack of adaptations in the map tasks. While the amount of time spent on the map task was not correlated with convergence for the current study, altering the task to ensure a minimum amount of time spent interacting may also aid convergence. A future study could achieve this by making dyads reproduce up to a certain amount of maps (*e.g.* 8 maps) until a certain time constraint is met. This would allow for a minimum amount of interaction time within the framework of the map task.

Even though PC was not noted for the dataset that included pre-task, during-task and post-task tokens, listeners may still detect it. Previous research on PC showed that acoustic and perceptual findings are not always compatible(Pardo et al., 2010; Pardo, Jordan, Mallari, Scanlon, & Lewandowski, 2013). These studies found that independent

listeners were able to detect changes to a dyads' speech after an interaction even though acoustic analyses of articulation rate and vowel formants either did not detect the same changes or noted the opposite pattern. To fully discount any adaptations due to interactions it would thus be important to conduct AXB listening tests. This would test the presence of PC in pre- and post-task dataset vs. during the maps. However, this is beyond the scope of this study and will be planned for future work.

Chapter 8: Role of Accent Imitation in Convergence

8.1 Introduction

A question has been raised in the course of this investigation regarding the role of explicit imitation in PC. Does the ability to imitate an accent affect a speaker's ability to converge to a partner? In other words, will speakers who are better able to approximate an accent converge to a greater extent than speakers who are not able to approximate an accent well? Giles' (1973) theory of accent mobility posits two levels of accents. The primary level involves a speaker's native language with the national dialect or standard variant lying at one end of a continuum and the speaker's regional dialect lying at the other. The secondary level involves accents that a speaker can mimic or approximate but which are not used on a regular basis. Transference is possible between the two levels. If a speaker is exposed to an unfamiliar accent (*e.g.* due to a move to a foreign country), information from the secondary level can be assimilated into the primary level. Other social or automatic theories of convergence do not predict a correlation between explicit accent imitation and convergence.

Studies that examine overt accent imitation do so with a focus either on better accent comprehension or on social aspects of interaction (Adank, Hagoort, & Bekkering, 2010; Adank, Stewart, Connell & Wood, 2013). Adank et al. (2010) found that speakers who repeated sentences in an unfamiliar accent while imitating that accent were better at speech comprehension in that unfamiliar accent than speakers who simply listened to the stimuli or repeated it in their own accent. With regard to the social aspects of interactions, Adank et al. (2013) reported that accent imitation can influence the way people evaluate others. Explicit imitation of a perceived accent improved listener ratings of speakers based on perceived power and competence and general social attractiveness. Both of

these studies suggest that explicit accent imitation has implications for how speakers interact and comprehend speech in that same accent.

One study has examined the role of explicit imitation on PC using AXB tasks (Pardo et al., 2010). They found that if providers were asked to imitate their partners, male providers converged. Convergence was noted more frequently in dyads where receivers were asked to imitate their partners. This study builds on this by examining the role of explicit imitation of an unfamiliar accent and its relationship to observed PC patterns.

Currently, there is a lack of research that explores the connection between pronunciation ability or talent in explicit accent imitation and the degree to which convergence occurs. In the current experiment, IS scores and centroid distance were used to approximate the degree of segmental and suprasegmental convergence. These measures were correlated with ratings of how well speakers imitated the Irish accent from an independent set of listeners to gauge the association between talent in accent imitation and PC.

It is possible that explicit imitation of an accent draws from the same pool of production and perception processes that give rise to implicit imitation such as convergence. If this is the case, speakers who demonstrate talent in imitating accents are more likely to adapt to the speech patterns of their interlocutor during a conversation. However if, as research suggests, convergence is the result of enduring changes to phonemic representations (mimesis) instead of imitation (Delvaux & Soquet, 2007; Nielsen, 2008), the prediction is that explicit accent imitation will not be correlated with any tendency to converge.

8.2 HYPOTHESES

Existing research suggesting that PC may be the result of mimesis (Delvaux & Soquet, 2007; Nielsen, 2008) leads to the prediction that neither segmental nor suprasegmental convergence scores will be correlated with imitation ratings. In other other words, talent in imitating a foreign accent will not predict one's ability to exhibit PC in either vowels or rhythm.

8.3 METHODOLOGY

8.3.1 Participants

8.3.1.2 Model speakers

One male and one female speaker served as the model speakers for this experiment. They were both native Irish English speakers who were born and brought up in Ireland. The female speaker's age was 25 at the time of recording; the male speaker's age is unknown. Both speakers had moved to the US after the age of 18.

8.3.1.2 Accent Raters

All 72 participants (36 male) from the native, mixed dialect and mixed language groups participated in the accent imitation task. A separate set of listeners scored the speakers' accent imitations. 11 listeners (5 male and 6 female, age unspecified) scored the male imitation data and 11 listeners (6 males and 5 females) scored female imitation data separately. All listeners were monolingual speakers of AE and were recruited through the Department of Psychology subject pool.

8.3.2 Design and procedure

8.3.2.1 Accent imitation

As the final post-task recording, participants from all three language conditions were presented this final paragraph:

> When the sunlight strikes raindrops in the air, they act as a prism and form a rainbow. The rainbow is a division of white light into many beautiful colors. These take the shape of a long round arch, with its path high above, and its two ends apparently beyond the horizon. There is, according to legend, a boiling pot of gold at one end. People look, but no one ever finds it. When a man looks for something beyond his reach, his friends say he is looking for the pot of gold at the end of the rainbow.

Once they had finished reading it, they were told that they would hear an Irish man or woman reading this paragraph. Their task was to imitate the accent of the speaker to the best of their abilities. All speakers heard either a male or female native Irish English speaker recite a sentence from this paragraph one at a time. They were then recorded imitating the model speaker. Male speakers imitated the male model speaker whereas female speakers imitated the female model speaker. Each speaker imitated a total of six sentences.

Sentences were segmented using a Praat script (Boersma & Weenik, 2003). A total of 216 sentences were obtained from male as well as female speakers. Another

script was used to normalize root mean square (RMS) of these sentences so that loudness was not a cue to accent rating.

8.3.2.1 Accent rating

A separate set of listeners was asked to score these imitations on a continuous scale using E-Prime. Each listener scored either the male or female data for a total of 216 sentences presented in randomized order. A listener heard the model speaker's version first followed by the imitated version for each of the 216 sentences. They were then asked to rate the sentences on a sliding scale from best imitation to worst imitation. Presentation of sentences was randomized for each rater to avoid any learning effects. This continuous scale was automatically converted to a rating score via E-Prime where '0' was the worst imitation and '600' was the best.

8.3.3 Acoustic and Statistical analysis

As mentioned above, E-Prime translated the sliding scale imitation rating scores to numerical values on a range of 0-600 (where 0 was the worst possible imitation ('worst imitation') and 600 was the best ('best imitation')). Each listener assigned a rating to each sentence. This resulted in each speaker getting 6 scores for six imitated sentences, which were then averaged across sentences and listeners to obtain a single explicit accent imitation rating score or simply, imitation score. This score was a measure of the perception of how well a speaker imitated the Irish accent.

Statistical analysis was conducted using R (R Core Team, 2012). As with the previous analyses, male and female data were separated and examined using correlation analyses. Correlation was tested for a speaker's vowels as well as rhythm. For vowels,

spec-IS was used as a measure of vowel-specific changes because it appears to be most sensitive to vowel adaptations. In order to assess the correlation between PC and accent imitation, each speaker was assigned a convergence score for vowels and for rhythm. For vowels, this score was the difference in mean spec-IS between post- and pre-task measurements. For rhythm, speakers were assigned convergence values based on the difference between the pre-task and post-task difference of the average rhythm centroid of the sentences from the recording paragraph and their partner's centroids (Table 8.1). Thus, a provider and a receiver that belonged to the same dyad got the same vowel and rhythm convergence scores. Correlation between the imitation score and each convergence score was examined using correlation (Pearson) in the Hmisc library in R.

	Formula
Vowel convergence score	Average spec-IS_{post} - Average spec-IS_{pre}
Rhythm convergence score	(EMS + $centroid_{provider,\ post}$ - EMS + $centroid_{receiver,\ post}$) - (EMS + $centroid_{provider,\ pre}$ - EMS + $centroid_{receiver,\ pre}$)

Table 8.1: Convergence scores (difference of scores) assigned to a speaker in each language condition.

8.4 RESULTS

A Pearson correlation between imitation and convergence scores revealed that imitating the Irish accent was not predictive of convergence in vowels or rhythm. The male and female correlation coefficients for vowel and rhythm are provided below in Table 8.2.

	Male coefficients	Female coefficients
Vowels	r = -0.12, n = 36, p = 0.4845	r = 0.06, n = 36, p = 0.7303
Rhythm	r = 0.1, n = 36, p = 0.1463	r = 0.25, n = 36, p = 0.5509

Table 8.2: Correlation coefficients for accent imitation and convergence in rhythm and vowels.

8.5 DISCUSSION

The current experiment was designed to examine the relationship between explicit imitation of an accent with the extent to which a speaker converges. A correlation between ability to imitate an accent and convergence in either vowels or rhythm was not noted in this experiment. This suggests that PC is not predicted by the ability to explicitly imitate. Lakin et al. (2003) suggest that mimesis could be implicated in PC. According to him, mimesis has social consequences and is used to aid communication by aligning speakers in mannerisms such as body language and speech. However, explicit imitation has no such implications and it does not predict a speaker's ability to demonstrate convergence in speech.

Giles' (1973) accent mobility theory posits that transference from the secondary level (which deals with explicit imitation of accents that are not used frequently) to the primary level (which deals with implicit imitation of ones native language and dialect) of the accent repertoire is possible. However, this transference is motivated by the need to communicate with speakers that use an unfamiliar accent or is triggered by social needs for example, wanting to sound more colloquial or refined. Giles (1973) called this downward or upward accent convergence, respectively. Nevertheless, even though transference between the two levels of an accent repertoire may be feasible, it is

necessitated by communicative and social motivations (such as wanting to sound more refined). Imitating an accent simply because one was asked to do so lacks any similar impetus. Therefore, an assimilation of information between the two levels of the accent repertoire does not predict that people who can imitate an accent well will converge to a greater degree than people who cannot imitate an accent.

Even if no correlation was found between conscious imitation and PC in the current experiment, accent mobility predicts that unconscious accent mimicry may be. Unconscious accent mimicry would be part of the secondary level of the accent repertoire and it might influence PC via transference from the secondary to the primary level. If it is the case, there is another possible manner in which to examine a connection between accent imitation (unconscious and conscious) and PC. A study could be designed that asks participants to imitate an accent and also to identify the feature or property that they were trying to imitate. These speakers then would interact with people who exhibit the same accent as the one the participants consciously imitated. In this manner, even for the people that cannot convincingly imitate an accent, perhaps they attend to the particular speech property they identified during the imitation task. It is possible that speakers will converge to this salient property during an interaction. Such a study may be more successful in exploring the correlation between different kinds of accent mimicry and PC.

Chapter 9: General Discussion and Conclusions

9.1 Overview

This general discussion draws all the results from previous chapters together into a combined interpretation. A discussion of the combined results of the native language, mixed dialect and mixed language groups is included in Section 9.2. This includes a discussion of the analyses used in the current study. This is followed by a discussion of future research and extensions. Discussion of the results from the map task itself and possible reasons for the lack of PC during interaction follows next. Finally, the chapter ends with conclusions and implications from this study.

Table 9.1, outlines the results from Chapters 4, 5 and 6. It summarizes the patterns of convergence and divergence with respect to effect of task noted in the aforementioned chapters. Those variables that did not show any effect of task are marked with '--' which indicates null results. The resultant general trends for each language condition are also provided in Table 9.2.

		Vowels									**Rhythm**
		Spec-IS						**Sys -IS**	**F1**	**F2**	**EMS + centroid**
Group	**Sex**	**/ɑ/**	**/æ/**	**/e/**	**/i/**	**/o/**	**/u/**				
NS_{AE}- NS_{AE}	Female	C						C	--	--	--
NS_{AE}- NS_{AE}	Male	--						--	--	--	C
NS_{AE}- NS_{IE}	Female	D	D	--	--	C	C	C	--	--	C*
NS_{AE}- NS_{IE}	Male	D						--	--	--	D
NS_{AE}- NN_{SP}	Female	--						--	--	--	--
NS_{AE}- NN_{SP}	Male	D	C	C	--	C	C	C	C	--	--

Table 9.1: Significant adaptation trends from Chapters 4, 5 and 6. C, D and '--' stand for convergence, divergence and undetermined respectively. C* specifies that two of the six speaker pairs demonstrated convergence.

9.2 DISCUSSION OF FINDINGS ACROSS LANGUAGE CONDITIONS

A combined discussion of all the results from Chapters 4, 5 and 6 is provided below. Recall that in order to manage the size of the models and keep acoustics separate, no direct comparisons between men and women or language conditions were made. Thus the discussion on sex effects and comparisons across all language conditions is largely speculative. Nevertheless, an attempt to bring those results together and speculate on combined trends and their possible interpretations is necessary to obtain a complete picture of PC within these groups.

9.2.1 Methodological considerations

9.2.1.1 VISC

Attempts at using VISC to detect vowel convergence were unsuccessful in the native language group. VISC is a measure of dynamic information in a vowel. It has been shown that dynamic measures are better at differentiating vowels than static measures (Hillenbrand, 2011). It is possible that PC is more easily detected by static vowel measures than dynamic ones because, even though static information at specific points in the vowel changes, overall the vowel retains its original identity by maintaining the dynamic information measured by VISC. For example, even if speakers converged in a particular vowel, phonetically, the vowel would have retained its identity. Consequently, the property that makes VISC better at vowel detection and differentiation may make it less sensitive in detecting PC.

As a result, attempts to examine VISC in the mixed dialect and mixed language group were not pursued. It is very possible that these results may have been different in the mixed dialect or mixed language groups, which had overlapping but not identical vowels. For example, in the mixed dialect group, female IE speakers raised their /ɔ/ vowels resulting in more /o/-like vowels whereas female AE speakers raised their /ɑ/ vowels to make them more /ɔ/-like. Considering that these resulted in a phonetic change in vowels, it is possible that VISC would have detected such an adaptation. Future research will explore this possibility in the mixed dialect and mixed language data.

9.2.1.2 Euclidean distance (ED)

Attempts at using ED in F1-F2 space to detect system-wide changes to vowel spaces also did not yield any significant results. If the orientation of a speaker's vowel

space changed with respect to their interlocutor's but the sum of the distance did not, ED would not note any overall modification. This suggests that this measure may not be sensitive enough to detect PC in cross-dialect and cross-language changes (however see Babel 2009a and 2010 for an example of this measure detecting convergence in cross-dialect interactions with an additional social manipulation).

9.2.1.3 IS

The IS measures provide a good starting point to quantify changes taking place in a dyad's vowel set. These measures are an improvement over the standard F1-F2 measurements because they are a measure of global (sys-IS) or pairwise (spec-IS) vowel similarity and are more sensitive than ED. The strength of the IS measures is based in their simplicity. These measures require a straightforward analysis because they consider the dyad as one unit. Furthermore, they quantify changes in F1 and F2 as one number. Both of these simplifications make interpretation of vowel space changes very straightforward. Because the dyad is considered together, individual linguistic background isn't a factor in the statistical analysis and need not be a factor until it is time to interpret the results.

9.2.1.4 Rhythm

EMS + centroid was capable of detecting differences in rhythm between speakers of L1 AE and L2 AE. However, as noted in Chapter 5, a consistent dialectal difference between IE and AE were not detected by EMS + centroid. One possible explanation for this is the small number of subjects analyzed. Another possibility is that this measure may be better at detecting inter-language rhythm differences but not intra-language

(dialectally variant) ones. EMS + centroid was capable of detecting rhythmic differences due to sentence structure/communicative role. Speakers, often receivers, whose sentence usage extended beyond statements to questions, hesitations and false starts were more likely to show a more variable rhythm when compared to their partners. Given that EMS + centroid is detecting sentence structure, speaking style and some linguistic differences in rhythmic structure, the need for further research in evaluating EMS + centroid (and other spectral measures) cannot be stressed enough. These measures provide a useful alternative to the traditional durational measures that quantify rhythm but require further attention and assessment in quantifying linguistic rhythm. It is possible that spectral and durational measures of rhythm are tracking different aspects of rhythm. While durational classifications have had some success tracking vocalic and consonantal variability, spectral measures are examining intensity changes or the concentration of energy in the amplitude envelope. Thus, durational and spectral measures may be quantifying linguistic rhythm in very different ways that are complementary. Regardless of how rhythm metrics track differences, listeners are capable of detecting cross-linguistic rhythm distinctions (Ramus & Mehler, 1999; Ramus et al., 2003; Ramus et al., 2000). The current findings suggest that speakers are not only capable of detecting linguistic rhythm distinctions but also adapting theirs in response to task and speaker demands.

9.2.2 General trends in vowels and rhythm

Based on Goldinger's (1997, 1998) findings that words in all interactions particularly low-frequency words lead to convergence, the predictions had been that speakers in all language conditions would show convergence particularly in the vowels that were marked within each group, specifically, /ɑ/ for the mixed dialect group and /ɑ,

æ/ for the mixed language group. Similarly, the prediction regarding rhythm was that speakers would show convergence in the native language group (because speakers shared the same language) and in the mixed language group (because L1 AE rhythm would be marked for Spanish speakers and L2 AE rhythm would be marked for AE speakers). The prediction in the mixed dialect group was a little more uncertain. Rhythm would be marked for IE speakers and AE speakers in the same manner as for the mixed language group predicting convergence. However, because these were proficient speakers of differing dialects of English, they might diverge to signal cultural distinctiveness. Despite a lack of a social manipulation involving group membership in the current setup, pilot data (Rao et al., 2011) had suggested that this was a possibility.

Dyads from all language conditions exhibited adaptations in their speech either in vowels or rhythm or both. The native language group showed convergence and it was noted in both vowels and rhythm. Within the mixed dialect group, both convergence and divergence were noted. Lastly, for the mixed language group, convergence was noted for just vowels.

Group	**Vowel**	**Rhythm**
Native language	C	C
Mixed Dialect	C & D	C & D
Mixed language	C	--

Table 9.2: General adaptation trends noted for each language condition. C, D and -- stand for convergence, divergence and undetermined respectively.

Separating results further by sex provides a more complex picture. PC trends for male and female speakers in the three language conditions are provided in Table 9.3. The trend is marked as convergence if either men or women within that group exhibited

convergence in vowels or rhythm. It can be seen that men and women exhibit differing adaptation strategies.

Group	Sex	Vowels	Rhythm
Native language	Female	C	--
Native language	Male	--	C
Mixed dialect	Female	C	C*
Mixed dialect	Male	D	D
Mixed language	Female	--	--
Mixed language	Male	C	--

Table 9.3: General adaptations noted in women and men in each group from Table 9.1. C, D and '--' stand for convergence, divergence and undetermined respectively. C* specifies that two of the six speaker pairs demonstrated convergence.

9.2.3 Differences due to sex of speaker and language background

There is an indication of sex differences across the language conditions. Though no direct comparisons between men and women were made, there was a trend of men converging in rhythm. Furthermore, these differences interacted with language background. With regards to rhythm, men in the linguistically close conditions (*i.e.* shared an L1) were more likely to adapt. Men in the native language group converged whereas men in the mixed dialect group diverged (again, presumably to maintain or increase social distance). Men in the mixed language group did not exhibit any rhythm adaptations. Additional research is necessary to identify the reason for the deviations noted in men and women. Nevertheless, these results suggest that convergence may be affected by both language background and speaker sex.

The findings from the current study are in line with Kim et al.'s (2011) theory of interlocutor language distance. Their study found that speakers who share the same

language and sub-dialect are more likely to converge than speakers who do not share the same language or sub-dialect. Dyads belonging to the native language group show patterns that are consistent with this theory. Both men and women showed convergence in either vowels or rhythm. Kim et al. (2011) did not detect convergence in the speech of native and non-native speakers, which they posit was due to interlocutors' desire to maintain social distance. Speakers in the mixed language group (different L1s) in the current study showed mixed results when compared to the Kim et al. (2011) study. Women, on the one hand, were consistent and showed no adaptations. Men, on the other hand, showed vowel-specific convergence (Table 9.1). This difference may be due to the different L1s of the non-native speakers in each study. Kim et al. (2011) used non-native speakers of AE who were L1 Korean speakers (in Illinois) whereas the current study used Spanish speakers (in Texas). It is possible that native AE speakers in this study had more experience with L1 SP speakers due to the large Hispanic and Latino community in Texas than those in Kim et al.'s (2011) did with L1 Korean speakers. Such previous experience might make it easier for a speaker living in Texas to converge with a L1 SP partner than for a speaker in Illinois (or even Texas) to converge with an L1 Korean speaker. Another possibility is that an unspecified social consideration caused men in the mixed language group to converge. Lastly, speakers from the mixed dialect group do not follow any trend noted by Kim et al. (2011) because proficient speakers of varying dialects such as IE or Singaporean or British English were not included in the study on linguistic distance. In this manner, the current study extends Kim et al.'s (2011) findings by examining acoustic changes in segmental and suprasegmental properties of speakers of two national varieties of a language.

With the exception of Krivokapic (2013), PC of rhythm has not been studied extensively. The pattern of rhythm convergence noted in the current study is best

explained by typological differences placed along the rhythm continuum (Figure 9.1, Dauer, 1983).

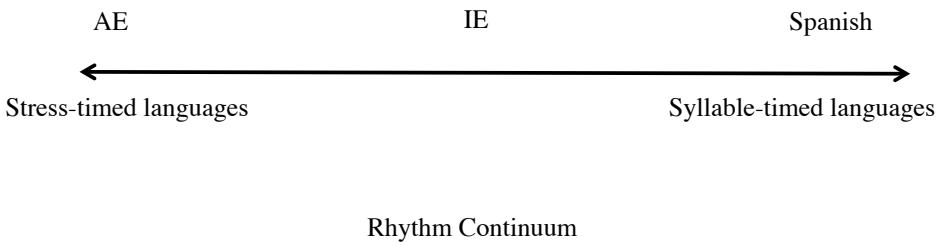


Figure 9.1: The rhythm continuum showing AE and Spanish at opposite ends with IE in the middle (after Dauer, 1983).

Rhythmic class may influence phonetic convergence in rhythm. Typologically, AE rhythm is considered stress-timed, SP rhythm is syllable-timed and IE rhythm is uncertain. While Krivokapic (2013) and Fuchs (2012) report syllable-timed rhythm, both these studies compare IE with another stress-timed dialect, either AE or BE not a syllable timed language such as SP. This only provides half the picture. If IE is mixed, it is very possible that comparing it to another syllable-timed language such as Spanish would reveal more stress-timed properties. Furthermore, Fuchs (2012) also found that some properties of syllable-timed languages, such as consonant cluster reduction, are less likely among educated IE speakers such as the participants in these studies. This lends support to the possibility that IE may be mixed with regards to syllable vs. stress timing. In the native language group, convergence was noted among speakers because L1 and national dialect (*i.e.* AE) were held constant across speakers. In the mixed dialect group, L1 was held constant but the dialect (*i.e.* AE and IE) was different. Both convergence and

divergence were noted among the speakers in this group. For the mixed language group, the two L1s, Spanish and AE, belong to typologically and perceptually different rhythmic classes that lie at opposite ends of the syllable-timed/stress-timed continuum (Figure 9.1). In this case, no adaptations were noted in rhythmic properties of speech. Thus, convergence is noted if the rhythmic properties of the interlocutors' language or dialect lie closer together along the rhythm scale. As this rhythmic correspondence decreases, possibility of adaptation decreases. Thus the difference between AE and Spanish may be too large to facilitate rhythm adaptations. On the other hand, IE rhythm may be altered to be more or less like AE because the two dialects share rhythmic properties, such as allowing vowel reduction and consonant clusters.

Previous studies showed that perceptual tasks were able to detect PC in the speech of both men and women in a pattern that did not necessarily line up with acoustic measures (Pardo et al., 2013). This suggests that when detecting PC, listeners may attend to more than one salient speech feature. Thus, during perceptual tests on the data from the current study, listeners may attend to information present in both vowels and rhythm (as well as other speech details) to detect adaptations in both men and women that pattern along the trends in Table 9.3. Thus, listeners may detect convergence in both men and women in the native language group. In fact, if, as Goldinger (1998) suggests, suprasegmental properties are more salient, listeners may detect more convergence in men's speech than women's as noted by Pardo (2006), Pardo et al., (2010) and Lewandowski (2012).

9.2.4 Role differences

The results also revealed an effect of role on vowels and rhythm though these differences were not consistent across language conditions. This is in line with studies by Kim et al. (2011) and Pardo (2006). The current findings are different from Pardo et al. (2010), which noted role-based differences based on which speaker role was instructed to imitate their partner. Instructing receivers to imitate resulted in more speaker pairs converging than the opposite condition of providers imitating their partner. Role-based adaptations may be viewed as speech modifications aimed at enhancing the interaction. They denote differences in speech patterns arising from the role assigned to interlocutors in a dyad. These variations may simply be due to the differences in the amount of speech each role produced. Specifically for rhythm, they may be a result of the influence of rhythmic variability from intonation patterns used during the task. While they are not a direct indicator of PC, these role-based differences do reveal an implicit awareness of 'responsibility' on the speaker's part. Interlocutors must have been aware of their role within the context of the map task to maintain vowel or rhythm differences throughout the task. This suggests that speakers altered their speech keeping task requirements in mind.

9.2.5 Theoretical implications

The findings from the current study cannot be interpreted within a purely dichotomous automatic or social framework. Even though social considerations other than linguistic distance were not explicitly manipulated in this study, the variations noted in convergence and divergence patterns within each group implicate social factors playing a role in PC. Goldinger's (1997) prediction that marked features of speech (with low frequency of occurrence) would be more susceptible to convergence is directly

contradicted in these findings. Instead, the patterns noted here suggest that marked features diverged presumably to denote out-group membership. For instance, in the case of the mixed dialect group, women diverged in /æ/ and /ɑ/ and converged in /u/ and /o/, while men diverged in vowels and rhythm. /ɑ/ is a marked vowel for this group and divergence in this vowel as well as in general male speech patterns suggests that social considerations may have been at play during the interactions. These speakers may have attempted to maintain distinctiveness of marked aspects of their speech to highlight their dialectal and cultural identity. Another instance of vowel-specific adaptation was noted in the men from the mixed language group. These speakers also showed divergence in one of the two vowels that were considered marked, /ɑ/ (the other vowel, /æ/ showed convergence). Such varying patterns of convergence and divergence in a group's speech patterns raise the possibility that all speech features need not converge to facilitate a conversation, increase task intelligibility and success likelihood. Speakers may demonstrate divergence in one or more marked speech components to signal membership in a group distinct from their partner's while simultaneously exhibiting convergence in other marked and non-marked speech components to facilitate communication. Consequently, an automatic theory of PC cannot account for this pattern of adaptations. Conversely, social theories of PC claim that speakers are able to adapt their speech patterns in order to manipulate a social factor such as identity or an interlocutor's attitude towards the speaker. Accordingly, a speaker is intentionally adapting his or her speech patterns to a partner. The only social factor that can account for the adaptation pattern noted in the current study is the sociolinguistic language manipulation. However, speakers were not told the purpose of the study until all recordings were obtained, suggesting that adaptations were a result of social considerations that were not conscious on the speaker's part. These findings provide support for Babel (2009a) and

Lewandowski's (2012) proposition that a combined theory that incorporates aspects of both automatic and social factors would be more appropriate to explain PC.

9.3 LIMITATIONS AND EXTENSIONS

Although this study showed that speakers from differing language backgrounds are capable of altering their vowels and rhythm, there are some notable limitations. The current study used a relatively small number of subjects, 6 dyads per sex per condition to yield a total of 72 speakers. Due to the time consuming nature of obtaining and analyzing the data, this is typical of PC research. For instance, Lewandowski (2012) used 20 non-native speakers in her study and Babel (2009a) used 150 participants across six conditions. Further research with a larger sample size in each condition may help to elucidate whether these trends are robust across speakers. Another limitation of the current study is that no direct comparisons were made between men and women and across language conditions. Further research that compares these factors directly in a larger-scale study is necessary to provide clearer insights into sex-based factors affecting PC.

EMS + centroid detected adaptations in speakers from varying language backgrounds. EMS + centroid tracks the center of mass of the rhythm spectrum whereas EMS and APS track peaks to detect syllable distribution and prominence and other melodic information that the amplitude spectrum provides (Liss et al., 2010; Rao & Smiljanic, 2011; Tilsen and Johnson, 2008). The presence or absence of energy in the higher frequencies of the amplitude envelope lowers or raises the centroid (or the most prominent peak). As a result, both typological differences in linguistic rhythm as well as different speaking styles (clear or citation or conversational) can lead to centroid/peak

movement making it difficult to determine how rhythm PC was achieved. In some cases, it is possible that both speakers shifted to clear speech resulting in lowered centroids. This is seen in the native language group where all men lowered their centroids denoting convergence. However, these speakers may also have switched to a more synchronized AE rhythm pattern (*e.g.* adopting a interlocutor's more neutral intonation pattern with less fluctuations) that resulted in lowered centroids due to reduced prosodic variability. Currently, all spectral measures are incapable of making this distinction. In order to ascertain whether speaking style changes or rhythmic and prosodic changes are leading to rhythmic convergence or divergence, a more controlled study should compare varying speaking styles and communicative situations using EMS + centroid with the aim of differentiating between speech style adaptations and rhythm synchronization.

Additional research is also necessary to determine which one of the two IS measures is more informative. To demonstrate, in the mixed dialect group, sys-IS showed that men's vowels diverged. However, spec-IS did not show any vowel-specific adaptations suggesting that separating vowel-specific similarity scattered the combined effect of the conversation on vowels. In the case of the mixed language group, sys-IS indicated vowel systemic convergence in men's speech. However, the more vowel-specific spec-IS revealed both convergence and divergence. Further research is required to clarify which IS measure is more informative, sys-IS or spec-IS. Given the current findings, sys-IS and spec-IS may be complementary and best used in tandem. Sys-IS is better at detecting system-wide changes whereas spec-IS is more of a measure of per-vowel similarity. For the current study, both sys-IS and spec-IS were considered when evaluating vowel convergence. Another shortcoming of the measures involves the direction of change. Because provider-receiver pairs are collapsed, no indication of the extent to which each speaker converges is available.

Research by Pardo and her colleagues has found that listeners can detect speech adaptations despite differing results from acoustical measurements (Pardo et al., 2010; Pardo et al., 2013). They argue for conducting production and perception studies together to gain a deeper understanding of PC. Given this recommendation, one extension of this work would be to conduct perceptual tests on this production data. AXB tests such as those used by Kim et al. (2011) or Pardo (2006) might reveal which trends noted in this study are in fact attended to by listeners. Goldinger (1998) has suggested that suprasegmental aspects of speech are most susceptible to PC. This theory predicts that listeners will detect PC in the speech of men and women that mirror the findings on rhythm convergence from the current study. If instead the vowels are more informative for listeners, they will detect PC in the speech of both women and men that is consistent with the pattern noted via vowel convergence. Some of these changes were very small (*e.g.* spec-IS value for men in the mixed language group changed from 0.9969 pre-task to 0.9975 post-task). Therefore, listener judgments would be helpful in determining which changes are noticeable and noteworthy and which are not.

Aside from the language background, this study did not manipulate any other social identity and social distance. In fact, the map task was selected specifically to decrease any effects of social motivators (*e.g.* attitudes) of PC by providing a common goal of reproducing the map route to the dyads. However, studies such as the current one that found varying language background without other social manipulations led to both convergence and divergence suggests that PC is inseparable from social considerations. An extension of this study that specifically includes an estimate of a speaker's attitude towards their partner and his or her culture would be very informative. Babel (2009a) used the IAT to obtain such a measure of a speaker's attitude towards the model speaker's race, which could also be employed within the framework of this study.

This study demonstrates that both segmental and suprasegmental properties of speech are susceptible to PC. However, it does not address the reasons for convergence or the underlying mechanisms involved in PC. Studies that have attempted to identify the purpose of PC suggest that it may aid communicative efficiency, intelligibility and task success (Shepard et al. 2001; Nenkova et al. 2008). However, the cognitive mechanisms and neural pathways involved in PC remain unexplored. Future studies that focus on the cognitive and neural mechanisms involved in speech-specific adaptations are a necessary extension of what is currently known about PC.

9.4 CORRELATION BETWEEN CONVERGENCE AND TASK EFFICIENCY

Nenkova et al. (2008) reported that speaker entrainment or convergence aided task success in a common goal oriented video game. They also noted that speakers who entrained were more efficient communicators. In order to test the correlation between convergence and task efficiency (as an approximation of task success) in the current study, speakers' vowel and rhythm scores were used to calculate a correlation between convergence and task efficiency. Each speaker was assigned a score based on pre-task and post-task spec-IS score differences (these scores were previously used in the explicit imitation task in Chapter 8). Spec-IS was used as a measure of vowel convergence because it was the most informative and detailed of all the vowel measures. Spec-IS measures for all vowels were averaged before the difference between pre-task and post-task was calculated. The same was done with centroid differences as well. For example, the difference between the pre-task and post-task spec-IS score was assigned to both the provider and receiver of each dyad of the native language group (Table 8.1).

As a result, this measure was a difference of difference score. This same pair was also assigned the difference between the receiver and provider's pre-task and post-task centroids. A positive value in the vowel score indicated convergence whereas a negative value for rhythm indicated divergence. These scores were tested against the amount of time spent completing the map task using a Pearson chi-squared test of correlation. No correlation was found between map efficiency and convergence scores for both vowels and rhythm. Mean times spent on each map by female and male dyads are provided below in Tables 9.4 and 9.5.

	Map 1	**Map 2**	**Map 3**	**Map 4**
Native language	5.42	7.07	5.87	7.07
Mixed dialect	4.16	3.37	4.34	4.20
Mixed language	9.12	10.86	12.44	10.05

Table 9.4: Average time spent by women completing each map separated by language condition.

	Map 1	**Map 2**	**Map 3**	**Map 4**
Native language	6.00	5.70	5.22	5.26
Mixed dialect	4.39	4.03	4.12	3.95
Mixed language	6.85	6.50	6.80	7.15

Table 9.5: Average time spent by men completing each map separated by language condition.

9.5 CONCLUSIONS AND IMPLICATIONS

The current study set out to investigate the effect of interlocutor linguistic distance on PC, specifically in vowels and rhythm. It found evidence that language distance may affect how men and women converge in vowels and rhythm. Dyads that

were close (*i.e*. they shared the same L1 and national dialect) converged in vowels and rhythm; dyads that were intermediate (*i.e*. they shared an L1 but spoke differing national varieties) converged in some aspects but diverged in others and dyads that were far (*i.e*. didn't share an L1) showed slight vowel convergence and divergence. Additionally, dyads diverged in some marked aspects of speech while converging in others to presumably signal out-group membership.

These findings are consistent with previous ones that both men and women are capable of adjusting their speech in response to an interaction (Babel, 2009a, b; Pardo, 2006; Krivokapic, 2013; Kim et al., 2011). As expected, novel measures of vowel and rhythm convergence proposed in this study were capable of detecting task-induced, listener-oriented adaptations. Additionally, they also noted a task-induced but role-based adaptation that suggests speakers maintain speech distinctions based on the demands of their conversational role during a task.

There is some indication that men and women adapt to different aspects of speech with women being more susceptible to vowel PC and men to rhythm PC. Previous research has suggested that women are more likely to alter their long-term speech patterns (particularly vowels) because they are more willing to adopt novel speech patterns and attempt the prestige speech variety (Eckert, 1998; Labov, 2001a). It stands to reason then that they might lead short-term adaptations in vowels as well. This was noted most clearly in the native language group where women converged in vowels but the men did not.

Results from the rhythm analysis raise an interesting possibility. With the exception of two female dyads from the mixed dialect group, rhythm adaptations were noted mostly in men. This may indicate an idiosyncrasy of this set of subjects. Alternatively, it may suggest that men may also contribute to sound change via

suprasegmental adaptations. Delvaux and Soquet (2006) have suggested that PC may be a source of long-lasting sound change. However, most research on changes such as these has focused on segmental features leaving suprasegmental features largely unexplored. One exception is research conducted by Coggshall (2009) who showed that since the 1970s, prosodic rhythm in English of the Lumbee people is being altered to denote group membership suggesting that, like segmental features of speech, rhythm is also subject to more long-term permanent changes. If short-term change feeds more permanent change in vowels, perhaps that is also the case for changes in rhythm. As noted in this study, men are demonstrating rhythm adaptations on a short-term scale; perhaps as with short-term changes, men lead long-term sound change in rhythm. Additional research into short-term and long-term rhythm change is necessary to confirm this speculation.

In the context of human cognition, PC may be the result of the same processes that aid and guide shared intentionality. Tomasello and colleagues (Tomasello, Carpenter, Call, Behne, & Moll, 2005) claim humans are unique in their ability to engage in collaborative acts with a common goal in mind. This shared intentionality allows for an exchange of experiences and information with other humans and may form the basis for cultural learning. In some cases, humans share information to the extent that it can be called altruistic because it only benefits the listener not the speaker (*e.g.* informing someone that their shoelace is undone) (Warneken & Tomasello, 2006; Tomasello, 2009). Language then is critical in directing this joint attention to a point of mutual interest. Tomasello et al., (2005) point out that language is a means to achieve but not necessarily the reason for shared intentionality. Furthermore, because these skills are demonstrated by pre-verbal children (and to a smaller extent by non-human primates such as chimpanzees), they posit that language and shared intentionality may derive from the same underlying processes and abilities.

Within this framework, seemingly disparate tasks used to evaluate PC such as diapix, map and synchronized reading tasks become to some extent equivalent as tasks that involve shared intentionality and a mutual exchange of information. PC then can be viewed as one strategy to aid joint attention and ensure task efficiency and success by improving intelligibility. Given that the current study demonstrates convergence, divergence and maintenance within the same dyads, this suggests that shared intentionality does not demand complete agreement at the phonetic level. It can be affected by social considerations such as racial bias, identity construction and group membership and agreement or disagreement with the interlocutor (Babel 2009a, Bourhis and Giles, 1977). Speakers are free to converge in aspects that are unmarked to achieve communicative goals while diverging in marked properties of speech to indicate individual, cultural and social distinctiveness.

Appendix A: relevant definitions

SPECTRUM

A spectrum is a Fourier analysis of a sound wave. In this process, a complex waveform is decomposed into its simpler component waveforms. The resultant spectrum is a graph of the component frequencies of these simpler waveforms plotted against their amplitudes. Spectra can be short-term or long-term. Short-term spectra are the result of analysis conducted on a specific time point of the waveform whereas long-term spectra are the result of analysis on a longer section of the waveform.

Component frequencies of a complex periodic speech wave are labeled as f, 2f, 3f … where f is the lowest component frequency and is called the fundamental frequency or f0. All subsequent frequencies are whole multiples of f0 and are called harmonics. In a spectrum, frequencies are represented graphically on the x-axis (in Hz) whereas amplitude is represented in the y-axis (in dB). An example of a spectrum is provided below in Figure D.1.

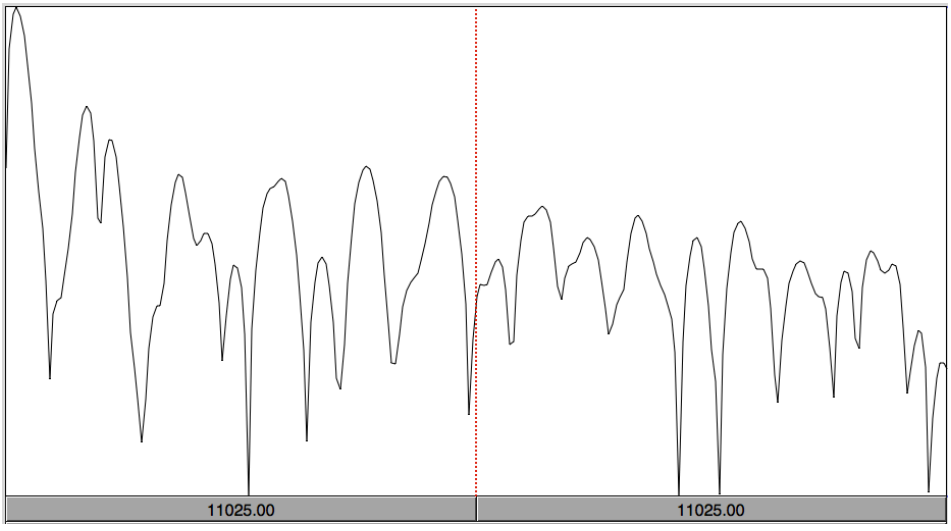


Figure D.1: Spectrum of a steady-state /i/ vowel as produced by an American male.

SPECTROGRAM

A sound wave is usually plotted as amplitude vs. time (waveform). However, it can also be plotted as frequency vs. time. A spectrogram is a plot of the intensity changes, as indicated by changes in darkness, for each frequency of a waveform as time progresses. All speech sounds yield characteristic spectrograms and show regularities that have been exploited in speech analysis. An example of a waveform and spectrogram is provided below in Figure D.2.

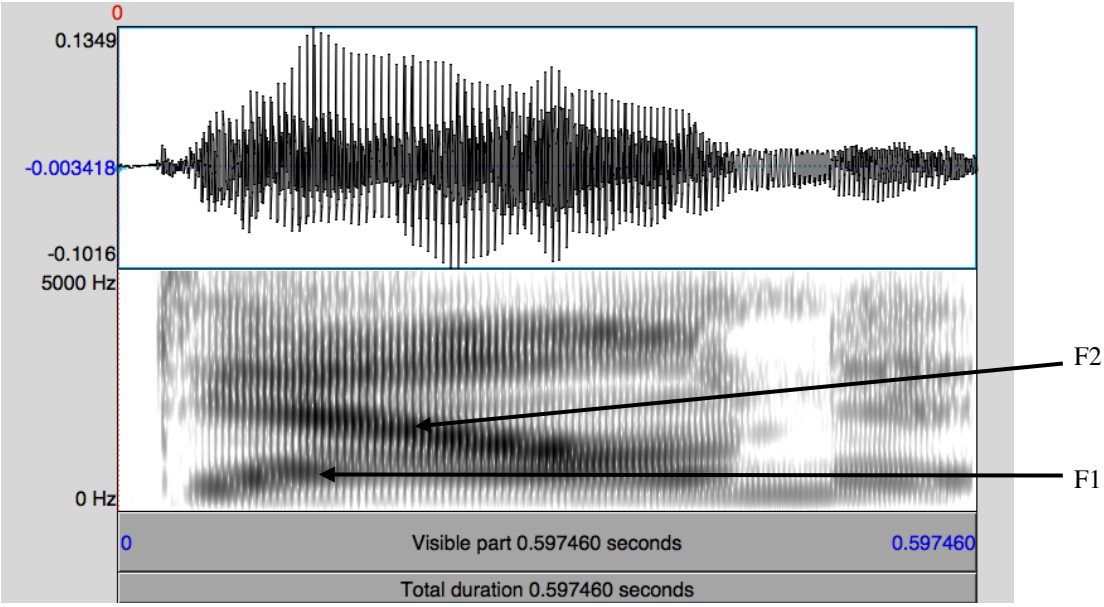


Figure D.2: Waveform (top) and spectrogram (bottom) representation of the word 'down' (produced by an American female). First (F1) and second (F2) formant frequencies are indicated with arrows.

FORMANTS

In a spectrogram, intensity is plotted on the z-axis, yielding bands of concentrated energy that vary systematically over time. These bands signify the resonances in the vocal tract and are called formants (regions labeled F1 and F2 in Figure 1.2 above). For vowel analysis, the three bands at the lowest frequencies (called the first (F1), second (F2) and third (F3) formants respectively) are used most frequently to differentiate among vowel categories. The formants arise from different vocal tract configurations, *i.e.* articulations. For example, F1 in vowels is related to vowel height. A higher F1 denotes a lower vowel with more open jaw.

VOWEL SPACE

The vowel inventory of a language can be represented as a plot of F1 vs. F2 (Peterson & Barney, 1952). This graph is called a vowel space. In addition to a language's vowel inventory being referred to as a vowel space, a person's individual and idiosyncratic vowel mappings can also be referred to as his or her vowel space. For the current study, each speaker's vowel space will be delineated by six peripheral vowels found in AE: /æ, ɑ, o, u, i, e/. Additional details are provided in the methodology in Chapter 3.

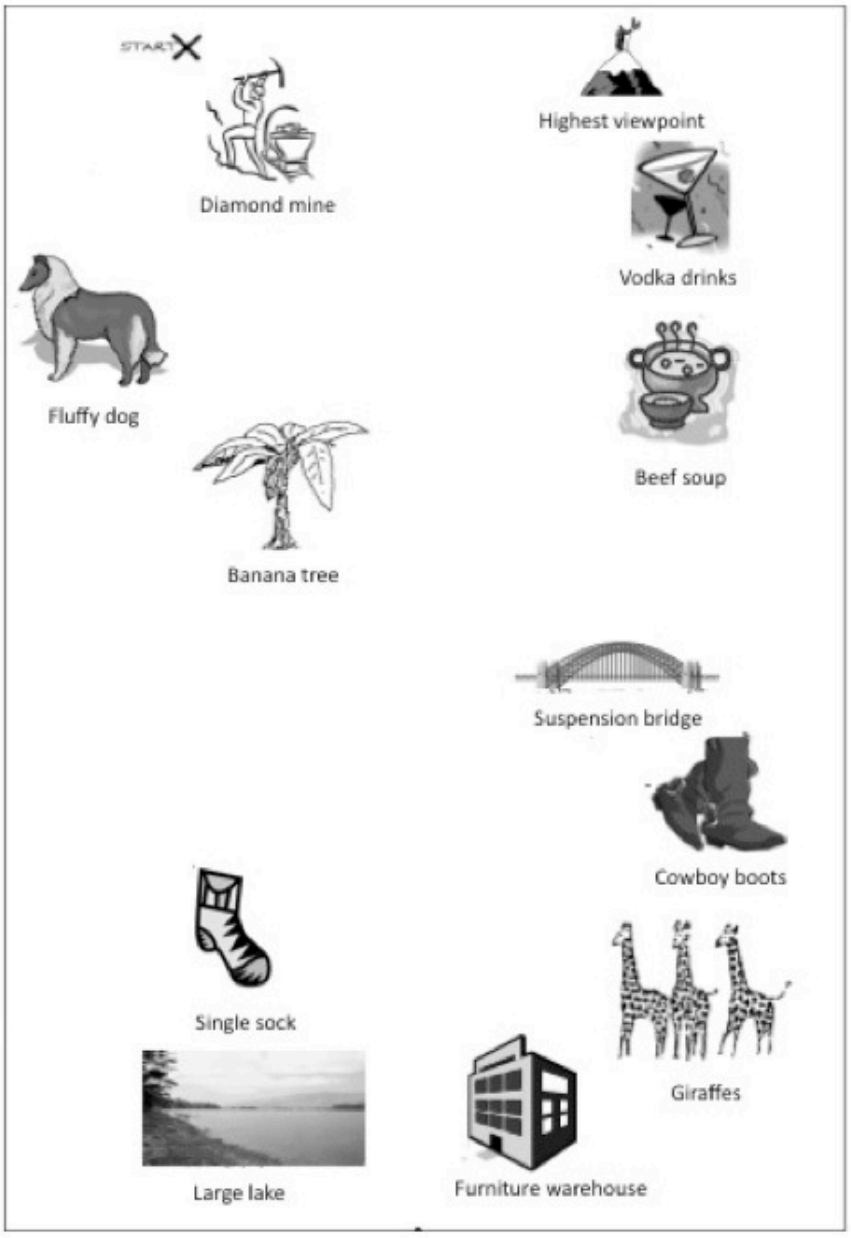
RHYTHM

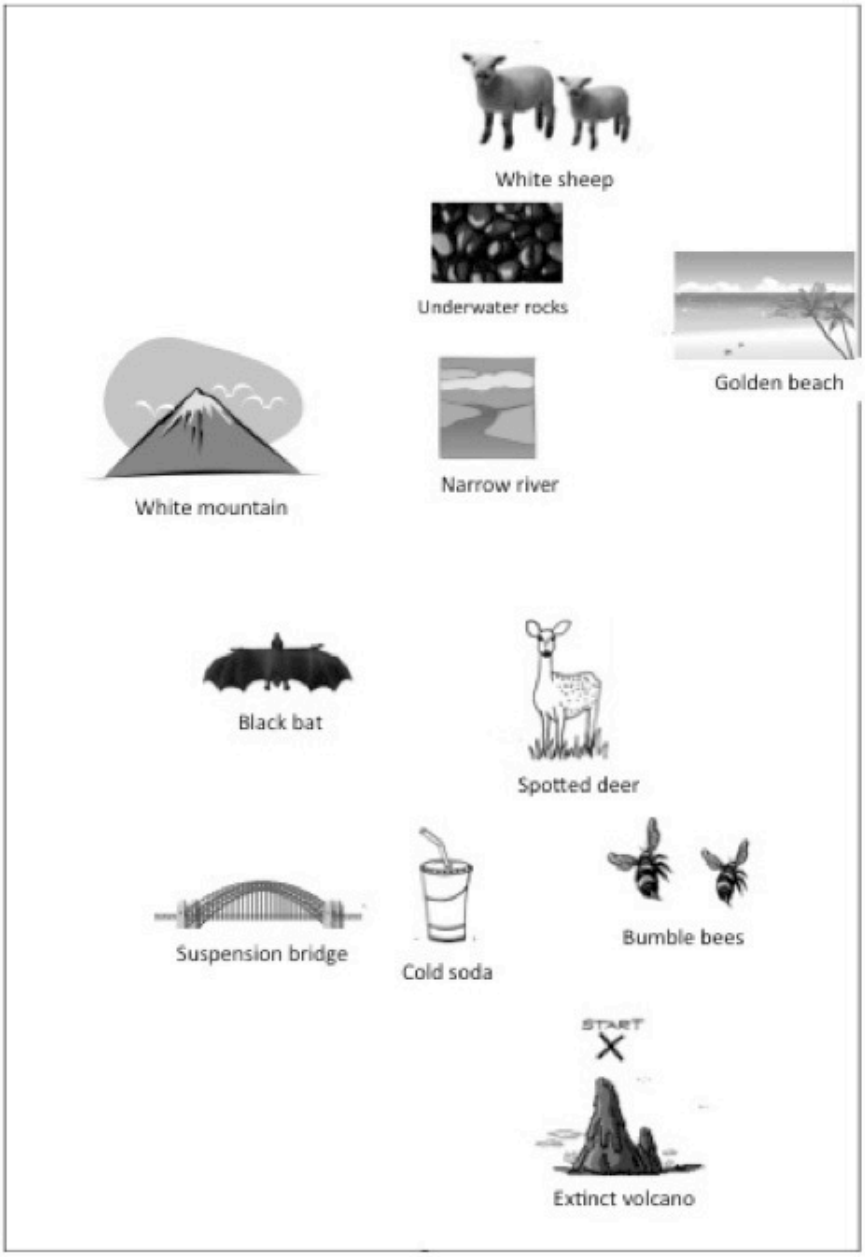
Languages vary in the patterns in which syllables are stressed and unstressed. The traditional view of rhythm posits that a language can be syllable-, stress-, mora- timed or mixed; although, this view has been called into question (Ramus et al., 2003; Dauer, 1983). Initial perceptual impressions of syllable timing have been described as the

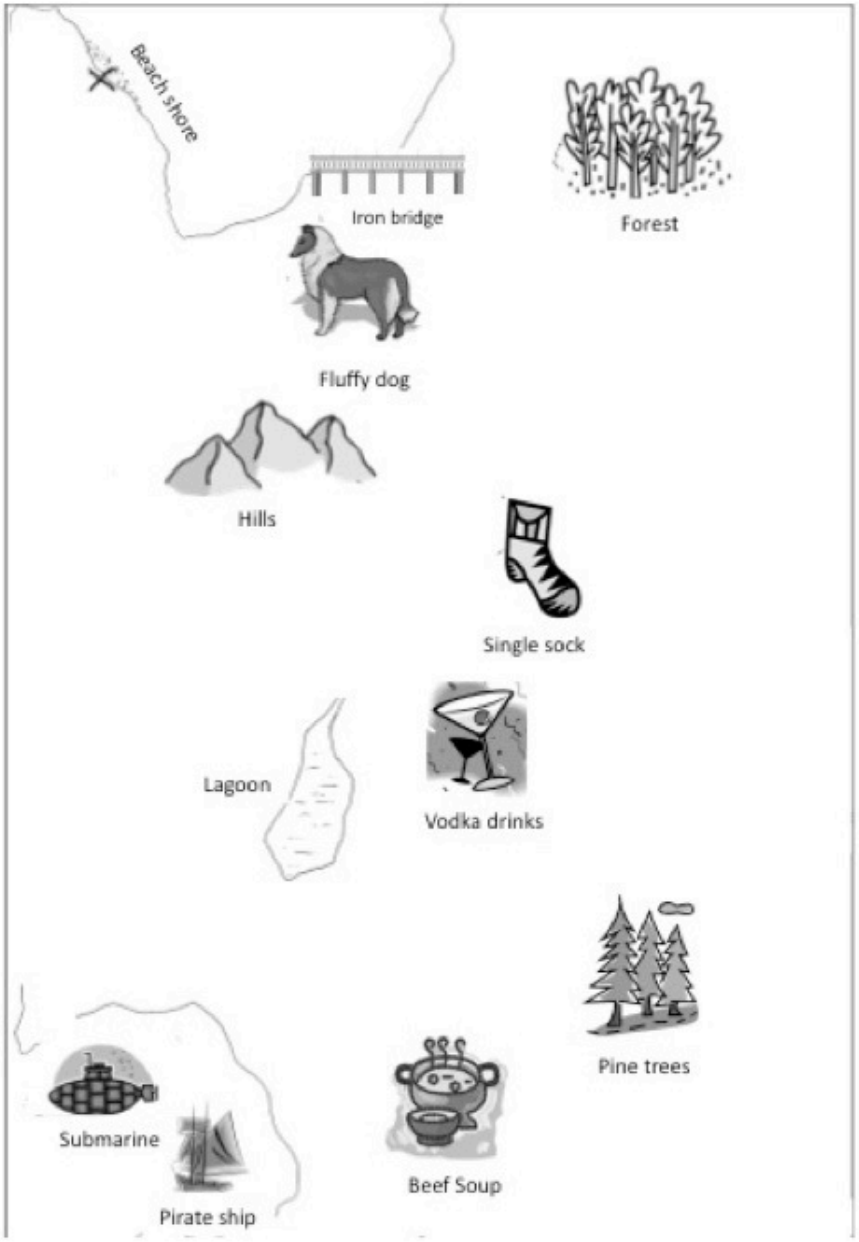
rhythm of a machine gun whereas stressed timing has been described as the rhythm of Morse code (Pike, 1946; Abercrombie, 1965). Syllable-timed languages tend to have isochronous syllables or syllables with equal durations (*e.g.* Spanish) whereas stress-timed languages tend to have isochronous inter-stress intervals (*e.g.* English) arising from more variable syllable structure (consonant clusters) and vowel reduction. A more recent definition of rhythm has been proposed by Patel (2008), which states that rhythm is the systematic patterning of sound in terms of timing, accent and grouping (p. 96).

Appendix B: maps

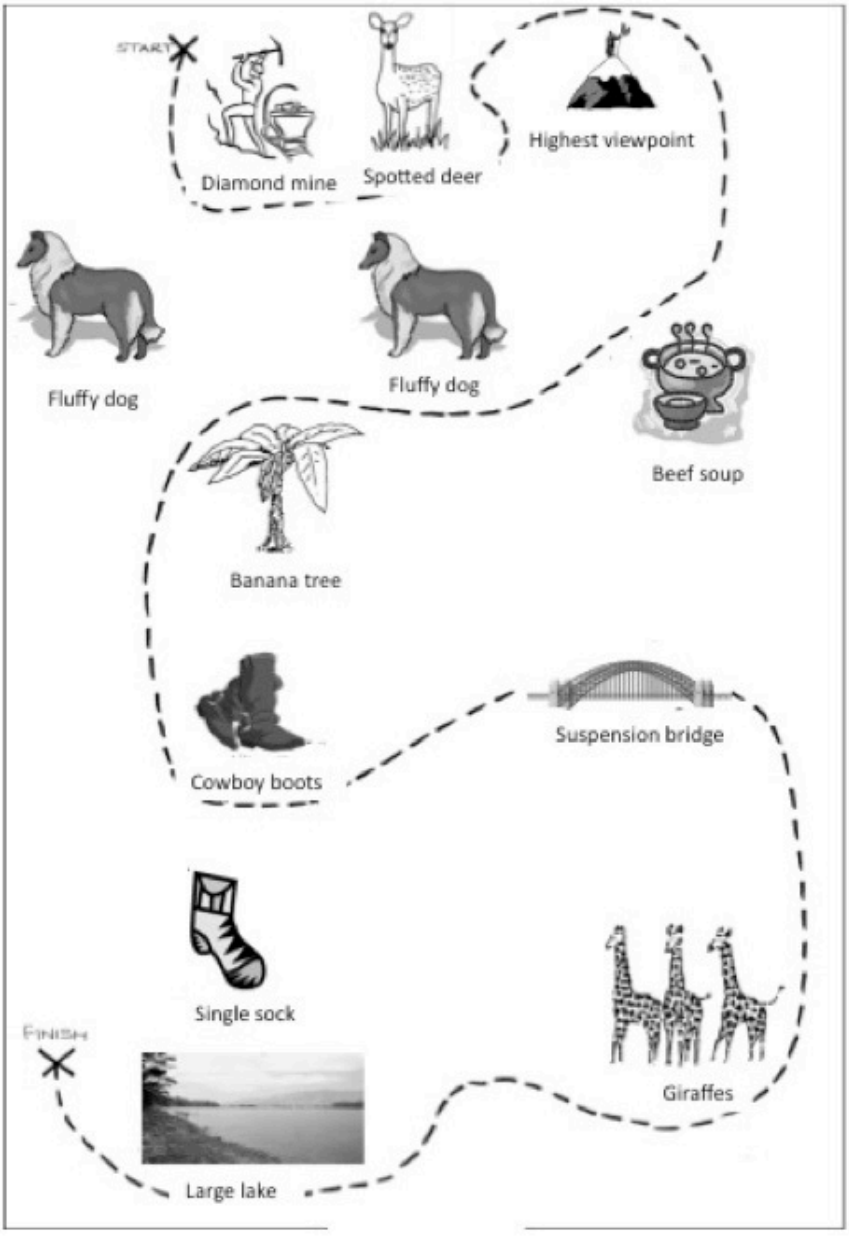
Receiver maps:


START
Diamond mine
Highest viewpoint
Vodka drinks
Fluffy dog
Beef soup
Banana tree
Suspension bridge
Cowboy boots
Single sock
Giraffes
Large lake
Furniture warehouse

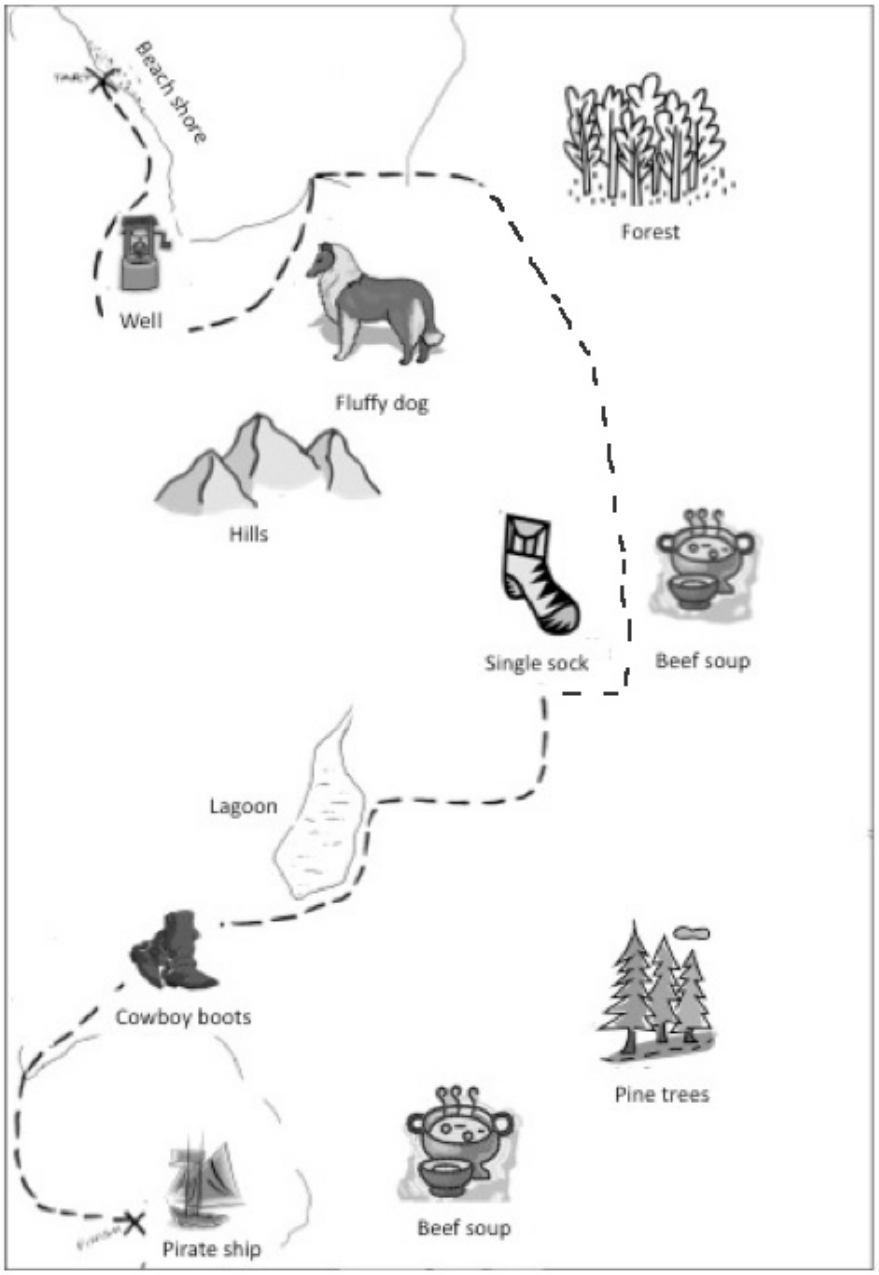

White sheep
Underwater rocks
Golden beach
White mountain
Narrow river
Black bat
Spotted deer
Suspension bridge
Cold soda
Bumble bees
START
Extinct volcano


Beach shore
Iron bridge
Forest
Fluffy dog
Hills
Single sock
Lagoon
Vodka drinks
Pine trees
Submarine
Pirate ship
Beef Soup

Provider maps:


START
Diamond mine
Spotted deer
Highest viewpoint
Fluffy dog
Fluffy dog
Beef soup
Banana tree
Cowboy boots
Suspension bridge
Single sock
Giraffes
FINISH
Large lake


Golden beach
White sheep
Golden beach
Narrow river
FINISH
White mountain
Black bat
Sleeping cat
Bumble bees
Suspension bridge
START
Cold soda
Extinct volcano


Beach shore
Forest
Well
Fluffy dog
Hills
Single sock
Beef soup
Lagoon
Cowboy boots
Pine trees
Beef soup
Pirate ship

Appendix C: modified LEAP-Q (Language Experience and Proficiency Questionnaire)

You may decline to answer any these questions

ID:
Today's date
Age
Date of birth
Male / Female
Ethnicity:
Race:

(1) Please list all the languages you know in the order of dominance:

Order	1	2	3
Language			

(2) Please list all the languages you know in order of acquisition (your native language first)

Order	1	2	3
Language			

(3) Please list what percentage of the time you are *currently* an *on average* exposed to each language. *(percentages should add up to 100%)*

Language	1	2	3
Percentage			

(4) When choosing to read a text available in all your languages, in what percentage of the cases would you choose to read it in each of your languages? Assume that the original was written in another language, which is unknown to you. *(percentages should add up to 100%)*

Language	1	2	3
Percentage			

(5) When choosing a language to speak with a person who is equally fluent in all your languages, what percentage of time would you choose to speak each language? Please report percent of total time. *(percentages should add up to 100%)*

Language	1	2	3
Percentage			

(6) Please list all the places that you have lived in for 2 years or longer.

	Country	Town	State/Province	Age during that period (eg. 0-12)
1				
2				
3				
4				

(8) Please name the cultures with which you identify. On a scale from zero to ten, please rate the extent to which you identify with each culture. (Examples of possible cultures include US-American, Chinese, Jewish-Orthodox, etc.) (0 = no identification; 10 = complete identification):

Culture	
Rate	

(9) How many years of formal education do you have?
Please check your highest education level (or the approximate U.S. equivalent to a degree obtained in another country):

O Less than High School	O Some College	O Masters
O High School	O College	O PhD/MD/JD
O Professional Training	O Some Graduate	O Other:

(10) Date of immigration to the United States, if applicable:
If you have ever lived in another country, please provide name of country and dates of residence:

(11) Have you ever had a:
O Vision problem O Hearing impairment O Language disability O Learning disability
(Check all applicable).
If yes, please explain (including any corrections):

(a) On a scale from zero to ten, please select your level of proficiency in speaking, understanding, and reading each language (0 = none; 10 = perfect):

Language	1	2	3
Speaking			
Understand spoken language			
Reading			

(b) On a scale from zero to ten, please select how much the following factors contributed to you learning each language (0 = not a contributor; 10 = most important contributor):

Language	1	2	3
Interacting with friends			
Interacting with family			
Language tapes/self-instruction			
Watching TV			
Reading			
Listening to the radio			

(c) Please rate to what extent you are currently exposed to each language in the following contexts (0 = never; 10 = always):

Language	1	2	3
Interacting with friends			
Interacting with family			
Listening to radio/music			
Reading			
Watching TV			
Language-lab/self-instruction			

(d) In your perception, how much of a foreign accent do you have in English (0 = none; 10 = pervasive)?

(e) Please rate how frequently others identify you as a non-native speaker of English based on your accent (0 = never; 10 = always):

If you are a speaker of American English:

> (f1) In your perception, how much of a regional accent do you have in English (0 = none; 10 = pervasive)?
>
> (f2) Please indicate which cultural or geographical area this regional accent is associated with

If you are a speaker of Spanish or Indian English:

> (g1) Please rate how frequently others identify you as accented speaker in English (0 = never; 10 = always)?

(g2) Please indicate which cultural or geographical area this regional accent is associated with

Appendix D: significant and non-significant statistics

PRE-TASK/POST-TASK VOWEL AND RHYTHM ANALYSES:

Native language group (NS_{AE}-NS_{AE})

Female dyads, F1:

	df	F-value	p-value
Vowel	5, 25	527.90	0.00*
Task	1, 5	0.00	0.99
Role	1, 60	14.77	0.00*
Vowel:Task	5, 25	0.70	0.62
Vowel:Role	5, 60	0.69	0.63
Task:Role	1, 60	0.27	0.60
Vowel:Task:Role	5, 60	0.21	0.95

Female dyads, F2:

	df	F-value	p-value
Vowel	5, 25	378.10	0.00*
Task	1, 5	5.29	0.06
Role	1, 60	0.34	0.56
Vowel:Task	5, 25	2.69	0.04
Vowel:Role	5, 60	3.52	0.007*
Task:Role	1, 60	1.86	0.17
Vowel:Task:Role	5, 60	0.32	0.89

Male dyads, F1:

	df	F-value	p
Vowel	5, 25	32.08	0.00*
Task	1, 5	0.47	0.52
Role	1, 60	12.57	0.00*
Vowel:Task	5, 25	0.86	0.51
Vowel:Role	5, 60	0.77	0.57
Task:Role	1, 60	0.35	0.55
Vowel:Task:Role	5, 60	0.13	0.98

Male dyads, F2:

	df	F-value	p-value
Vowel	5, 25	180.20	0.00*
Task	1, 5	0.15	0.71
Role	1, 60	7.32	0.00*
Vowel:Task	5, 25	0.16	0.97
Vowel:Role	5, 60	1.93	0.10
Task:Role	1, 60	0.01	0.91
Vowel:Task:Role	5, 60	0.09	0.99

Female dyads, rhythm:

	df	F-value	p-value
Role	1, 7	0.67	0.44
Speaker	5, 70	8.67	0.00*
Task	1, 84	0.07	0.79
Role:Speaker	7, 70	4.04	0.002*
Role: Task	1, 84	0.02	0.89
Speaker:Task	5, 84	0.87	0.50
Speaker:Task:Role	5, 84	1.15	0.33

Male dyads, rhythm:

	df	F-value	p-value
Role	1, 7	0.86	0.38
Speaker	5, 70	1.63	0.16
Task	1, 84	7.74	0.006*
Role:Speaker	7, 70	0.43	0.82
Role: Task	1, 84	1.17	0.28
Speaker:Task	5, 84	2.73	0.02*
Speaker:Task:Role	5, 84	1.08	0.37

Mixed dialect group (NS_{AE}-NS_{IE})

Female dyads, F1:

	df	F-value	p-value
Vowel	5, 10	587.10	0.00*
Task	1, 2	1.33	0.36
Role	1, 72	9.58	0.002*
Dialect	1, 72	4.28	0.04
Vowel:Task	5, 10	2.59	0.09
Vowel:Dialect	5, 72	1.83	0.11
Task:Dialect	1, 72	1.25	0.26
Vowel:Role	5, 72	0.42	0.83
Task:Role	1, 72	0.14	0.70
Dialect:Role	1, 72	3.39	0.06
Vowel:Task:Dialect	5, 72	0.03	0.99
Vowel:Task:Role	5, 72	0.03	0.99
Vowel:Dialect:Role	5, 72	1.30	0.27
Task:Dialect:Role	1, 72	0.03	0.84
Vowel:Task:Dialect:Role	5, 72	0.07	0.99

Female dyads, F2:

	df	F-value	p-value
Vowel	5, 10	168.90	0.00*
Task	1, 2	0.29	0.64
Role	1, 72	8.51	0.004*
Dialect	1, 72	46.38	0.00*
Vowel:Task	5, 10	0.62	0.68
Vowel:Dialect	5, 72	13.88	0.00*
Task:Dialect	1, 72	0.00	0.95
Vowel:Role	5, 72	1.63	0.16
Task:Role	1, 72	1.95	0.16
Dialect:Role	1, 72	0.01	0.89
Vowel:Task:Dialect	5, 72	0.22	0.95
Vowel:Task:Role	5, 72	1.06	0.39
Vowel:Dialect:Role	5, 72	2.63	0.03
Task:Dialect:Role	1, 72	0.52	0.47
Vowel:Task:Dialect:Role	5, 72	0.19	0.96

Male dyads, F1:

	df	F-value	p-value
Vowel	5, 10	494.80	0.00*
Task	1, 2	0.20	0.69
Role	1, 72	9.71	0.002*
Dialect	1, 72	2.11	0.15
Vowel:Task	5, 10	1.63	0.23
Vowel:Dialect	5, 72	4.73	0.00*
Task:Dialect	1, 72	0.35	0.55
Vowel:Role	5, 72	2.95	0.01*
Task:Role	1, 72	0.74	0.39
Dialect:Role	1, 72	10.59	0.001*
Vowel:Task:Dialect	5, 72	0.15	0.97
Vowel:Task:Role	5, 72	0.18	0.96
Vowel:Dialect:Role	5, 72	2.85	0.021*
Task:Dialect:Role	1, 72	0.61	0.43
Vowel:Task:Dialect:Role	5, 72	0.08	0.99

Male dyads, F2:

	df	F-value	p-value
Vowel	5, 10	210.60	0.00*
Task	1, 2	0.42	0.58
Role	1, 72	7.89	0.006*
Dialect	1, 72	72.43	0.00*
Vowel:Task	5, 10	3.53	0.04
Vowel:Dialect	5, 72	8.93	0.00*
Task:Dialect	1, 72	0.03	0.85
Vowel:Role	5, 72	1.19	0.32
Task:Role	1, 72	0.03	0.85
Dialect:Role	1, 72	9.42	0.003*
Vowel:Task:Dialect	5, 72	0.03	0.99
Vowel:Task:Role	5, 72	0.06	0.99
Vowel:Dialect:Role	5, 72	0.60	0.69
Task:Dialect:Role	1, 72	0.03	0.86
Vowel:Task:Dialect:Role	5, 72	0.01	0.99

Female dyads, rhythm:

	df	F-value	p-value
Dialect	1, 8	15.13	0.004*
Role	1, 16	3.17	0.09
Speaker	2, 128	7.74	0.00*
Task	1, 32	2.86	0.10
Dialect:Role	1, 16	1.15	0.29
Role:Task	1, 32	0.03	0.86
Dialect:Task	1, 32	1.13	0.29
Dialect:Speaker	2, 128	8.39	0.00*
Role:Speaker	2, 128	4.88	0.01*
Task:Speaker	2, 128	8.50	0.00*
Dialect:Role:Task	1, 32	0.21	0.64
Dialect:Role:Speaker	2, 128	6.15	0.002*
Dialect:Task:Speaker	2, 128	2.67	0.07
Role:Task:Speaker	2, 128	4.01	0.020*
Dialect:Role:Task:Speaker	2, 128	0.13	0.87

Male dyads, rhythm:

	df	F-value	p-value
Dialect	1, 8	10.39	0.01*
Role	1, 16	17.17	0.00*
Speaker	2, 128	2.33	0.10
Task	1, 32	4.76	0.03*
Dialect:Role	1, 16	0.67	0.42
Role:Task	1, 32	0.01	0.90
Dialect:Task	1, 32	0.00	0.99
Dialect:Speaker	2, 128	1.95	0.14
Role:Speaker	2, 128	11.86	0.00*
Task:Speaker	2, 128	1.62	0.20
Dialect:Role:Task	1, 32	1.12	0.29
Dialect:Role:Speaker	2, 128	1.02	0.36
Dialect:Task:Speaker	2, 128	3.41	0.03
Role:Task:Speaker	2, 128	0.34	0.71
Dialect:Role:Task:Speaker	2, 128	0.12	0.88

Mixed language group (NS_{AE}-NN_{SP})

Female dyads, F1:

	df	F-value	p-value
Vowel	5, 10	373.80	0.00*
Task	1, 2	2.09	0.28
Role	1, 72	2.63	0.11
Language	1, 72	18.14	0.00*
Vowel:Task	5, 10	1.19	0.37
Vowel:Language	5, 72	7.78	0.00*
Task:Language	1, 72	0.34	0.56
Vowel:Role	5, 72	4.08	0.002*
Task:Role	1, 72	0.77	0.38
Language:Role	1, 72	0.21	0.64
Vowel:Task:Language	5, 72	0.09	0.99
Vowel:Task:Role	5, 72	0.26	0.93
Vowel:Language:Role	5, 72	3.55	0.006*
Task:Language:Role	1, 72	0.06	0.80
Vowel:Task:Language:Role	5, 72	0.23	0.94

Female dyads, F2:

	df	F-value	p-value
Vowel	5, 10	204.80	0.00*
Task	1, 2	6.228	0.13
Role	1, 72	6.688	0.01*
Language	1, 72	8.141	0.005*
Vowel:Task	5, 10	0.086	0.99
Vowel:Language	5, 72	1.757	0.13
Task:Language	1, 72	1.379	0.24
Vowel:Role	5, 72	0.679	0.64
Task:Role	1, 72	0.906	0.34
Language:Role	1, 72	0.285	0.59
Vowel:Task:Language	5, 72	0.515	0.76
Vowel:Task:Role	5, 72	0.929	0.46
Vowel:Language:Role	5, 72	1.411	0.23
Task:Language:Role	1, 72	0.250	0.61
Vowel:Task:Language:Role	5, 72	0.095	0.99

Male dyads, F1:

	df	F-value	p-value
Vowel	5, 10	530.10	0.00*
Task	1, 2	47.07	0.02*
Role	1, 72	0.17	0.67
Language	1, 72	0.31	0.57
Vowel:Task	5, 10	3.02	0.06
Vowel:Language	5, 72	1.58	0.17
Task:Language	1, 72	0.53	0.46
Vowel:Role	5, 72	0.56	0.73
Task:Role	1, 72	0.00	0.95
Language:Role	1, 72	0.13	0.71
Vowel:Task:Language	5, 72	0.06	0.99
Vowel:Task:Role	5, 72	0.54	0.74
Vowel:Language:Role	5, 72	1.03	0.40
Task:Language:Role	1, 72	0.21	0.64
Vowel:Task:Language:Role	5, 72	1.30	0.27

Male dyads, F2:

	df	F-value	p-value
Vowel	5, 10	199.70	0.00*
Task	1, 2	4.81	0.15
Role	1, 72	2.42	0.12
Language	1, 72	15.75	0.00*
Vowel:Task	5, 10	0.90	0.51
Vowel:Language	5, 72	7.94	0.00*
Task:Language	1, 72	0.08	0.77
Vowel:Role	5, 72	1.11	0.36
Task:Role	1, 72	0.08	0.77
Language:Role	1, 72	0.86	0.35
Vowel:Task:Language	5, 72	0.08	0.99
Vowel:Task:Role	5, 72	0.42	0.83
Vowel:Language:Role	5, 72	0.11	0.98
Task:Language:Role	1, 72	1.52	0.22
Vowel:Task:Language:Role	5, 72	0.87	0.50

Female dyads, rhythm:

	df	F-value	p-value
Language	1, 10	0.06	0.81
Role	1, 20	2.61	0.12
Speaker	2, 160	0.48	0.61
Task	1, 40	0.59	0.44
Language:Role	1, 20	1.55	0.22
Role:Task	1, 40	0.78	0.38
Language:Task	1, 40	0.19	0.66
Language:Speaker	2, 160	0.99	0.37
Role:Speaker	2, 160	0.66	0.51
Task:Speaker	2, 160	1.45	0.23
Language:Role:Task	1, 40	0.01	0.91
Language:Role:Speaker	2, 160	7.88	0.00*
Language:Task:Speaker	2, 160	0.30	0.73
Role:Task:Speaker	2, 160	0.42	0.65
Language:Role:Task:Speaker	2, 160	1.94	0.14

Male dyads, rhythm:

	df	F-value	p-value
Language	1, 10	8.49	0.01*
Role	1, 20	0.50	0.48
Speaker	2, 160	7.74	0.00*
Task	1, 40	0.86	0.34
Language:Role	1, 20	6.01	0.02*
Role:Task	1, 40	1.75	0.19
Language:Task	1, 40	0.96	0.33
Language:Speaker	2, 160	0.15	0.85
Role:Speaker	2, 160	3.52	0.00*
Task:Speaker	2, 160	3.46	0.03
Language:Role:Task	1, 40	0.96	0.33
Language:Role:Speaker	2, 160	12.87	0.00*
Language:Task:Speaker	2, 160	3.26	0.04
Role:Task:Speaker	2, 160	0.04	0.96
Language:Role:Task:Speaker	2, 160	0.02	0.97

PRE-TASK, DURING-TASK AND POST-TASK RHYTHM ANALYSES:

Native language group (NS_{AE}-NS_{AE})

Female dyads:

	β-value	SE	t-value	p-value
Intercept	6.98	0.26	27.09	0.00*
Role	0.82	0.39	2.10	0.03
Task	0.08	0.08	0.93	0.34
Speaker	0.23	0.09	2.41	0.01*
Role:Task	-0.20	0.13	-1.49	0.13
Role:Speaker	-0.13	0.14	-0.93	0.35
Task:Speaker	-0.01	0.03	-0.34	0.73
Role:Task:Speaker	0.06	0.05	1.22	0.22

Male dyads:

	β-value	SE	t-value	p-value
Intercept	7.51	0.28	26.47	0.00
Role	0.06	0.44	0.14	0.88
Task	0.03	0.09	0.37	0.71
Speaker	0.25	0.09	2.60	0.01*
Role:Task	0.01	0.15	0.05	0.96
Role:Speaker	0.01	0.15	0.10	0.91
Task:Speaker	-0.03	0.03	-0.84	0.40
Role:Task:Speaker	0.01	0.05	0.17	0.86

Mixed dialect group (NS_{AE}-NS_{IE})

Female dyads:

	β-value	**SE**	**t-value**	**p-value**
Intercept	7.33	0.34	21.48	0.00
Dialect	0.50	0.49	1.01	0.31
Role	1.60	0.54	2.93	0.00*
Task	0.08	0.11	0.73	0.46
Speaker	0.10	0.28	0.35	0.72
Dialect:Role	-1.58	0.74	-2.11	0.03
Dialect:Task	-0.02	0.17	-0.11	0.90
Role:Task	-0.14	0.18	-0.79	0.42
Dialect:Speaker	0.18	0.39	0.46	0.64
Role:Speaker	-0.56	0.41	-1.33	0.18
Task:Speaker	0.06	0.09	0.64	0.52
Dialect:Role:Task	0.07	0.25	0.29	0.76
Dialect:Role:Speaker	0.39	0.59	0.65	0.51
Dialect:Task:Speaker	-0.16	0.13	-1.19	0.23
Role:Task:Speaker	-0.02	0.14	-0.17	0.86
Dialect:Role:Task:Speaker	0.10	0.20	0.53	0.59

Male dyads:

	β-value	**SE**	**t-value**	**p-value**
Intercept	7.59	0.45	16.62	0.00
Dialect	1.16	0.64	1.80	0.07
Role	0.76	0.63	1.20	0.23
Task	0.12	0.14	0.86	0.38
Speaker	-0.27	0.28	-0.95	0.34
Dialect:Role	-0.79	0.91	-0.86	0.38
Dialect:Task	-0.12	0.21	-0.59	0.54
Role:Task	0.09	0.20	0.47	0.63
Dialect:Speaker	-0.58	0.40	-1.46	0.14
Role:Speaker	0.39	0.40	0.95	0.33
Task:Speaker	-0.01	0.09	-0.12	0.90
Dialect:Role:Task	0.06	0.29	0.21	0.83
Dialect:Role:Speaker	0.29	0.57	0.51	0.60
Dialect:Task:Speaker	0.14	0.13	1.06	0.28
Role:Task:Speaker	-0.16	0.13	-1.25	0.20
Dialect:Role:Task:Speaker	-0.01	0.19	-0.06	0.95

Mixed language group (NS_{AE}-NN_{SP})

Female dyads:

	β-value	**SE**	**t-value**	**p-value**
Intercept	8.21	0.33	24.51	0.00
Dialect	0.32	0.25	1.29	0.19
Role	0.03	0.51	0.05	0.95
Task	-0.16	0.10	-1.54	0.12
Speaker	-0.35	0.25	-1.36	0.17
Dialect:Role	-0.29	0.37	-0.78	0.43
Dialect:Task	0.04	0.08	0.54	0.58
Role:Task	0.21	0.17	1.22	0.21
Dialect:Speaker	0.17	0.19	0.87	0.38
Role:Speaker	0.71	0.41	1.71	0.08
Task:Speaker	0.12	0.08	1.46	0.14
Dialect:Role:Task	-0.05	0.12	-0.43	0.66
Dialect:Role:Speaker	-0.48	0.29	-1.62	0.10
Dialect:Task:Speaker	-0.07	0.06	-1.04	0.29
Role:Task:Speaker	-0.15	0.13	-1.12	0.25
Dialect:Role:Task:Speaker	0.07	0.09	0.73	0.46

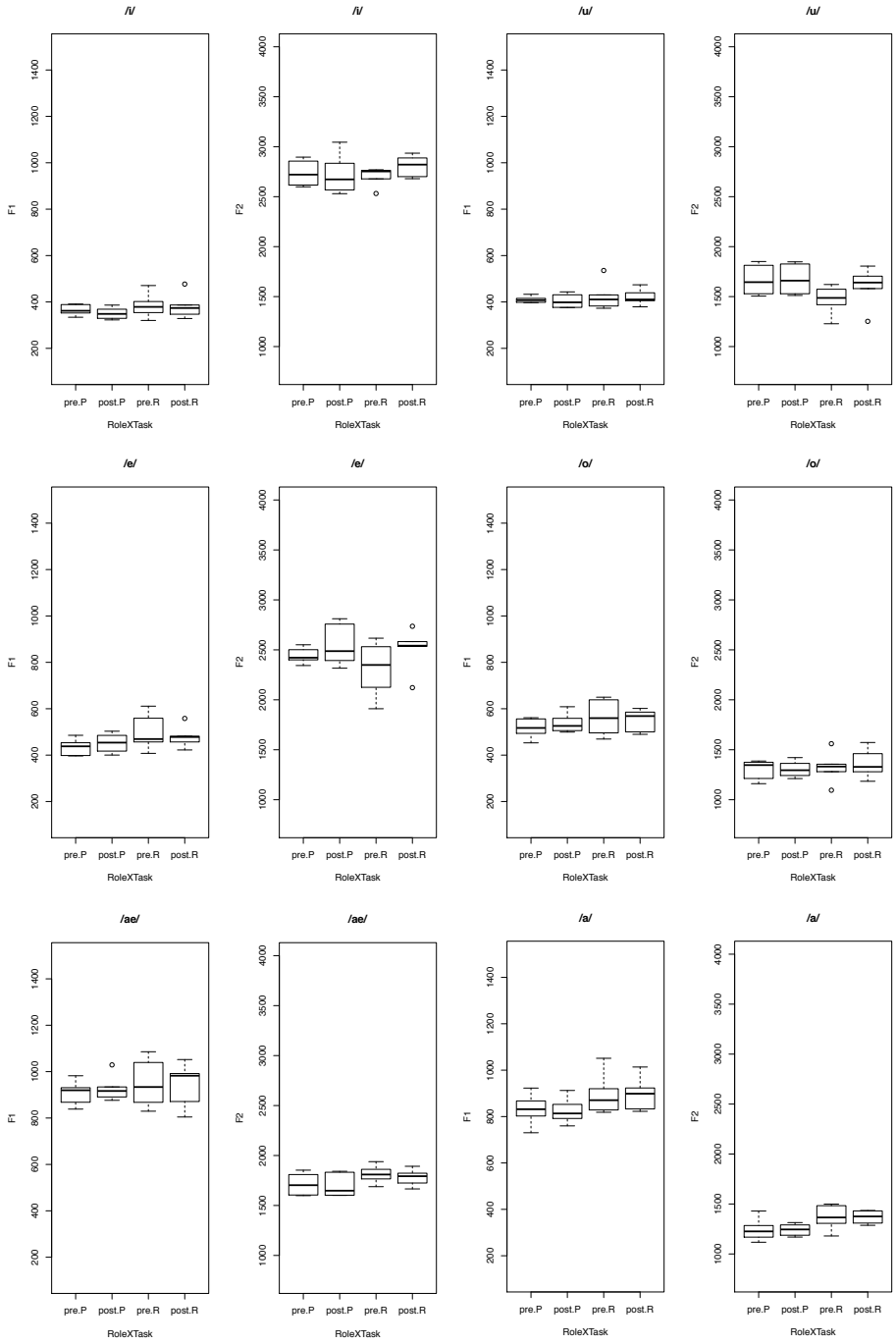
Male dyads:

	β-value	**SE**	**t-value**	**p-value**
Intercept	8.14	0.30	26.28	0.00
Dialect	-0.24	0.22	-1.08	0.27
Role	-0.06	0.48	-0.13	0.89
Task	-0.07	0.10	-0.74	0.45
Speaker	-0.64	0.25	-2.53	0.01*
Dialect:Role	-0.05	0.32	-0.15	0.87
Dialect:Task	0.07	0.07	0.96	0.33
Role:Task	0.28	0.16	1.71	0.08
Dialect:Speaker	0.23	0.17	1.34	0.17
Role:Speaker	0.72	0.39	1.83	0.06
Task:Speaker	0.04	0.08	0.55	0.57
Dialect:Role:Task	-0.07	0.11	-0.69	0.48
Dialect:Role:Speaker	-0.16	0.26	-0.63	0.52
Dialect:Task:Speaker	-0.02	0.05	-0.43	0.66
Role:Task:Speaker	-0.14	0.13	-1.10	0.26
Dialect:Role:Task:Speaker	-0.01	0.08	-0.08	0.93

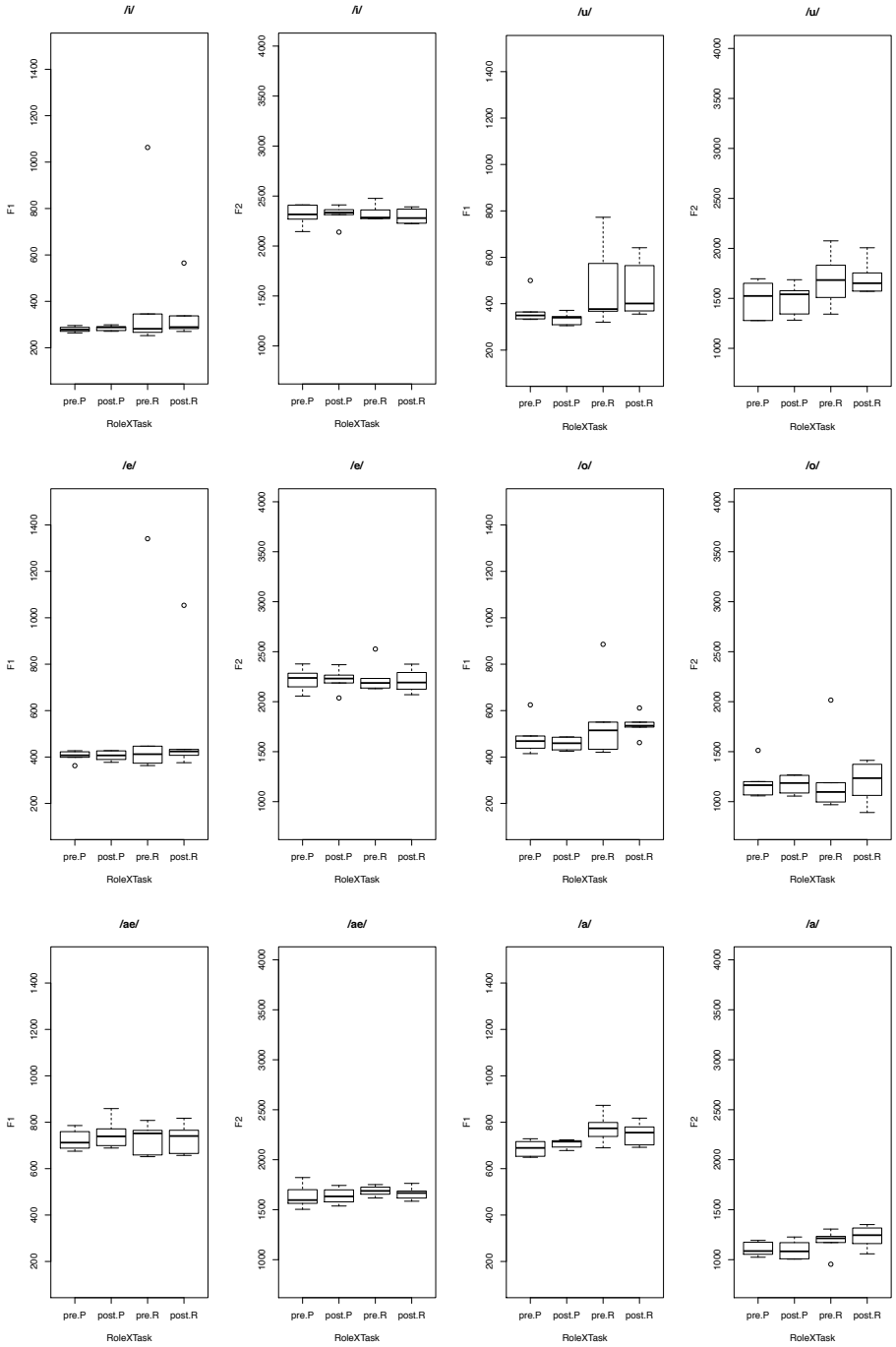
Appendix E: alternative descriptive plots of vowels

NATIVE LANGUAGE GROUP (NS_{AE}-NS_{AE})

Female dyads:

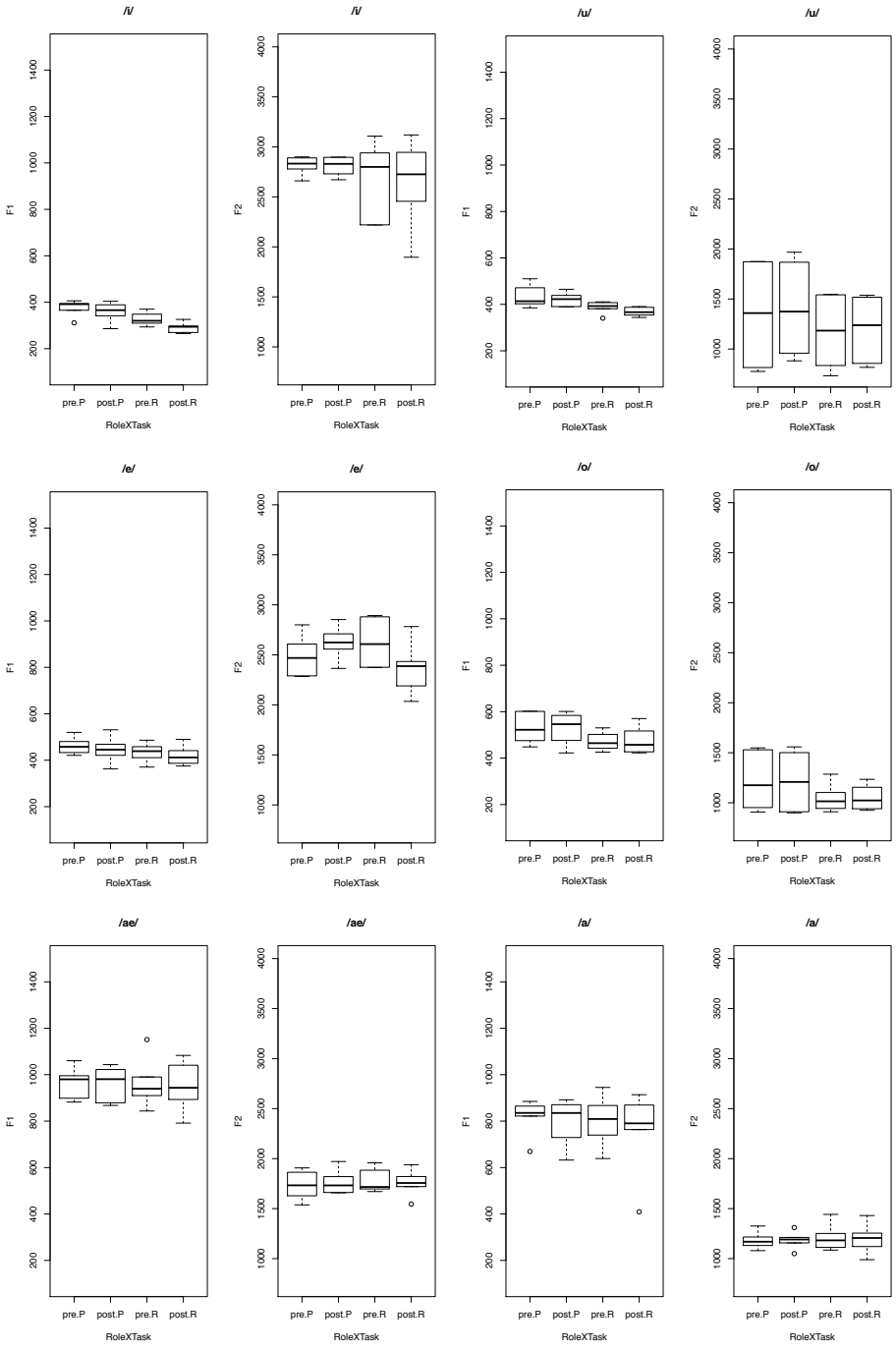


Male dyads:

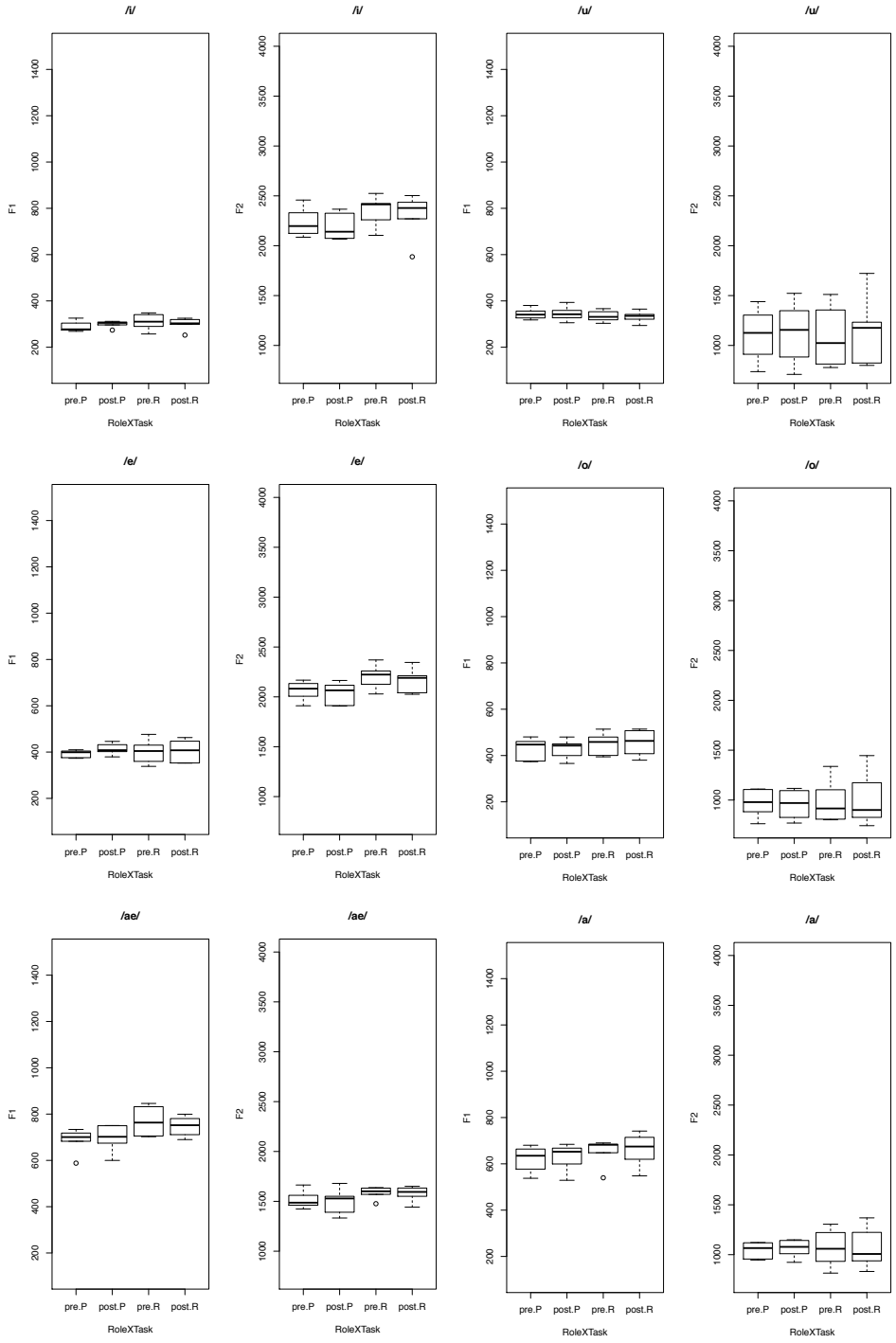


MIXED DIALECT GROUP (NS_{AE}-NS_{IE})

Female dyads:

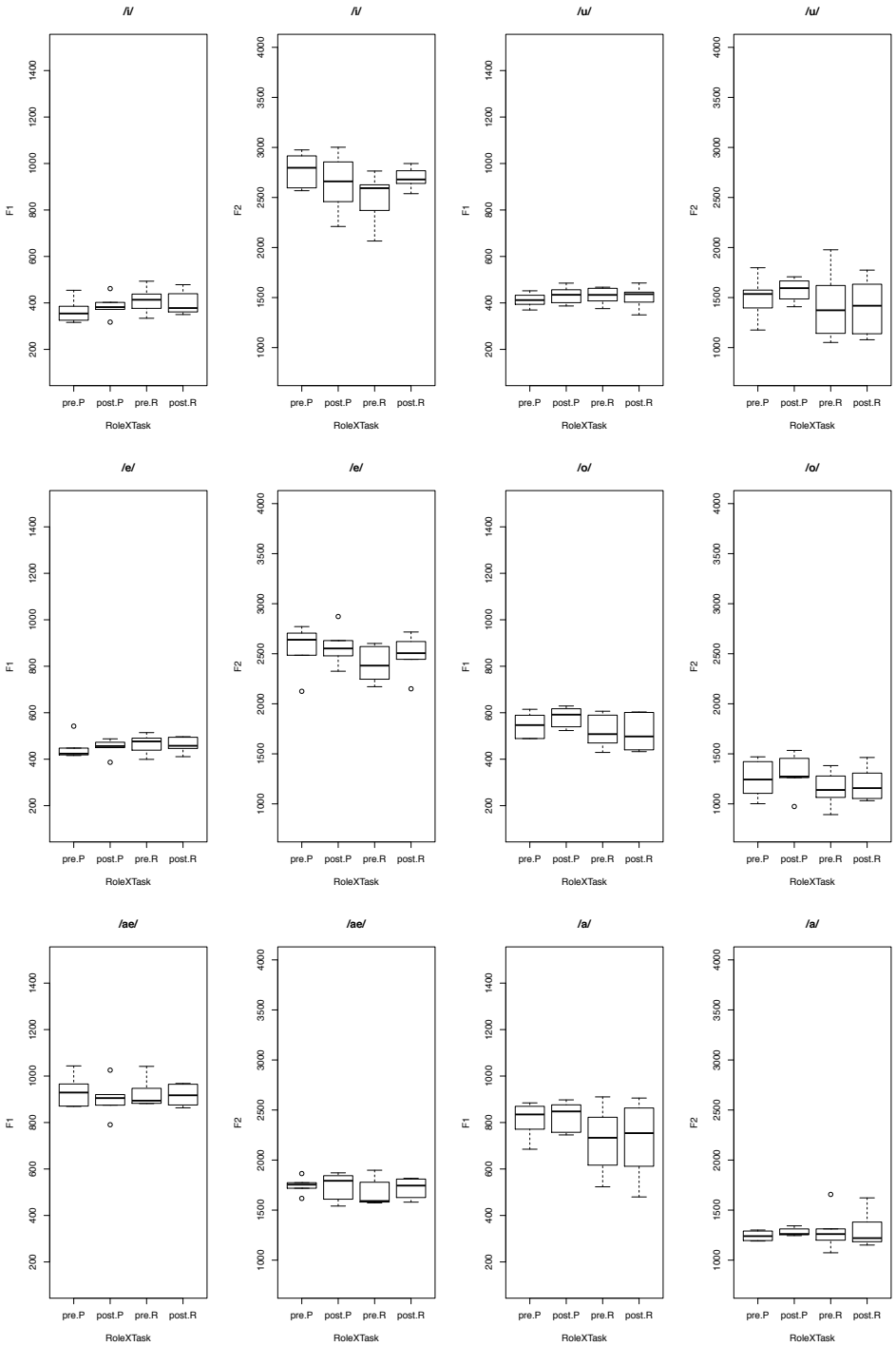


Male dyads:

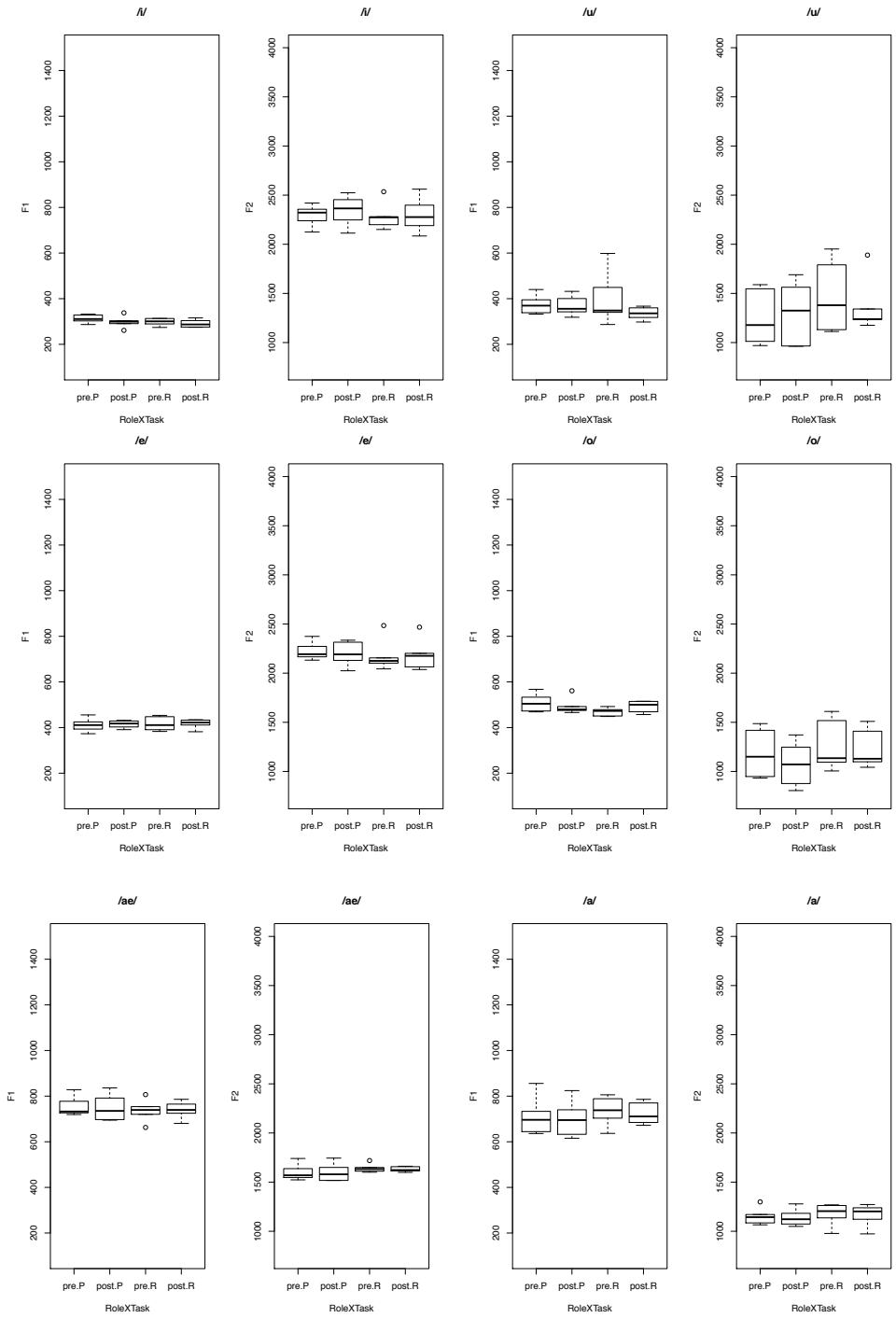


MIXED LANGUAGE GROUP (NS_{AE}-NN_{SP})

Female dyads:



Male dyads:



References

Abercrombie, D. (1965). *Studies in general phonetics*. Scotland, UK: Edinburgh University Press.

Anderson, A. H., Bader, M., Bard, E. G., Boyle, E., Doherty, G., Garrod, S., Isard, S., Kowtko, J., McAllister, J., Miller, J., Sotillo, C., Thompson, H. S., and Weinert, R. (1991). The HCRC Map Task corpus. Language & Speech, 34, 351-366.

Arvaniti, A. (2009). Rhythm, timing and the timing of rhythm. *Phonetica*, 66, 46-63.

Babel, M. E. (2009a). Phonetic and Social Selectivity in Speech Accommodation. Unpublished doctoral dissertation, University of California, Berkley.

Babel, M. E. (2010). Dialect convergence and divergence in New Zealand English. *Language in Society*, 39 (4), 437-456.

Baker, R. E., Bonnasse-Gahot, L., Van Engen, K., Baese-Berk, M. & Kim, M. (2009). Word Level Rhythm in Non-Native English. *Poster presentation at the Acoustical Society of America,* Portland, Oregon.

Beckman, M. E., Yoneyama, K., & Edwards, J. (2003). Language-specific and language-universal aspects of lingual obstruent productions in Japanese-acquiring children. *Journal of the Phonetic Society of Japan*, 7, 18-28.

Bloch, B. (1950). Studies in colloquial Japanese IV: Phonemics. Language, 26, 86-125.

Bloomfield, L. (1935). *Language*. London : Allen & Unwin.

Boersma, P. & Weenink, D. (2010). Praat: doing phonetics by computer [Computer program]. Version 5.1.31, retrieved 4 April 2010 from http://www.praat.org/

Carter, P. M. (2005). Quantifying rhythmic differences between Spanish, English, and Hispanic English. *In Proceedings of the 34th Symposium on Romance Languages*, 63–75.

Chartrand, T. L., Maddux, W., & Lakin, J. (2005). Beyond the perception-behavior link: The ubiquitous utility and motivational moderators of nonconscious mimicry. In R. Hassin, J. Uleman, & J.A. Bargh (Ed.). Unintended thought II: The new unconscious. New York: Oxford University Press.

Coggshall, E. (2008) The Prosodic Rhythm of Two Varieties of Native American English. *U. Penn Working Papers in Linguistics*, Volume 14(2).

Cummins, F. (2003). Entraining speech with speech and metronomes. *Cadernos de Estudos Linguisticos*, 70, 43-55.

Dauer, R. M. (1983). Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics*, 11, 51-62.

Dellwo, V. (2006). Rhythm and Speech Rate: A Variation Coefficient for deltaC. In Karnowski, P., Szigeti, I. (Eds.) Language and Language Processing (pp. 231-241). Frankfurt, Germany: Peter Lang.

Delvaux, V. & Soquet, A. 2007. The Influence of Ambient Speech on Adult Speech Productions through Unintentional Imitation. *Phonetica*, 64, 145-173.

Diehl, R. L., Lotto, A. J. & Holt, L. L. (2004). Speech perception. *Annual Review of Psychology*, 55, 149-179.

Drullman, R., Festen, J. M. & Plomp, R. (1994). Effect of temporal envelope smearing on speech reception. *Journal of Acoustical Society of America*, 95, 1053-1064.

Edlund, J., Heldner, M., & Gustafson, J. (2005). Utterance segmentation and turn-taking in spoken dialogue systems. *Sprachtechnologie, mobile Kommunikation und linguistische Ressourcen, 57*6-587.

Eckert, P. (1998). Adolescent social structure and the spread of linguistic change. *Language in Society*, *17*, 183–207.

Flege, J. E., Mackay, I. R. A. & Piske, T. (2002). Assessing bilingual dominance. *Applied Psycholinguistics*. 23, 567–598.

Frota, S. & Vigário, M. (2001). On the correlates of rhythmic distinctions: The European/Brazilian Portuguese case. *Probus*, 13, 247-275.

Fuchs, R. (2012). A duration-based account of speech rhythm in Indian English Phonology of Indian English Results. *poster presentation at Laboratory Phonology (LabPhon) 13, Stuttgart, Germany*.

Giles H. (1973). Accent mobility: A model and some data. *Anthropological Linguistics*, 15, 87-105.

Goldinger, S. D. (1997). Perception and production in an episodic lexicon. In Talker Variability in Speech Processing. Johnson, K. and Mullennix J. W. (Ed.). 33–66. Academic Press: San Diego.

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, 105(2), 251–279.

Grabe E & Low E. L. (2003) Durational variability in speech and the rhythm class hypothesis. In: Papers in laboratory phonology (7), 515-546.

Hay, J. F., Sato, M., Coren, A. E., Moran, C.L. & Diehl, R. L. (2006). Enhanced contrast for vowels in utterance focus: A cross-language study. *Journal of the Acoustical Society of America*, 119, 3022-3033.

Hillenbrand, J. M. (2011). Static and dynamic approaches to understanding vowel perception. In G.S. Morrison and P.F. Assmann (Ed.). Vowel Inherent Spectral Change, Heidelberg: Springer-Verlag.

Hillenbrand, J.M., Getty, L.A., Clark, M.J., and Wheeler, K. (1995). Acoustic characteristics of American English vowels. *Journal of the Acoustical Society of America*, 97, 3099-3111.

Jakobson, R., Fant, G. & Halle, M. (1952). *Preliminaries to Speech Analysis. The distinctive features and their correlates*. MIT Press, Cambridge, Massachusetts.

Junqua, J. C. (1993). The Lombard reflex and its role on human listeners and automatic speech recognizers. *Journal of the Acoustical Society of America*, 93, 510–524.

Kendall, R. (2002). Musical timbre beyond a single note, II: Interactions of pitch chroma and spectral centroid. *Proceedings of the 7th International Conference on Music Perception and Cognition, Sydney: Causal Productions, Adelaide, 596-599*.

Kim, M., Horton, W. S., & Bradlow, A. R. (2011). Phonetic convergence in spontaneous conversations as a function of interlocutor language distance. *Laboratory Phonology*, *2 (1)*(2011), 125–156.

Krivokapic, J. (2010). Prosodic interaction between speakers of American and British English. *Poster presented at the 159th Meeting of the Acoustical Society of America Baltimore, Maryland, April 2010*.

Krivokapic, J. (2013). Rhythm and convergence between speakers of American and Indian English. *Laboratory Phonology*, 4(1), 39–65.

Krull, D. & Engstrand, O. (2003). Speech rhythm – intention or consequence? Cross-language observations on the hyper/hypo dimension. *Reports from the Department of Phonetics*, University of Umeå (PHONUM) 9, 133-136.

Kuhl, P. K., Andruski, J. E., Chistovich, L., Chistovich, I. , Kozhevnikova, E., Sundberg, U., and Lacerda, F. (1997). Cross language analysis of phonetic units in language addressed to infants. *Science* 227.684–6.

Labov, W. (1963). *The social motivation of a sound change*. Word 19, 273–309.

Labov, W. (2001a). *Principles of Linguistic Change, Volume I: Internal Factors*. Blackwell Publishing, Cambridge, MA.

Labov, W. (2001b). *Principles of linguistic change, Volume II: Social Factors*. Blackwell Publishing, Cambridge, MA.

Labov, W. (2002). Driving Forces in Linguistic Change. *International Conference on Korean Linguistics*.

Labov, W., Ash, S., & Boberg, C. (2006). *The Atlas of North American English: Phonetics, Phonology, and Sound Change : a Multimedia Reference Tool. Supplement* (p. 318). Walter de Gruyter.

Ladefoged, P (2001). *Vowels and consonants: an introduction to the sounds of languages*. Blackwell Publishing, Cambridge, MA.

Lakin, J., Jefferis, V., Cheng, C., and Chartrand T. (2003). The chameleon effect as social glue: evidence for the evolutionary significance of nonconscious mimicry. *Journal of Nonverbal Behavior* 27(3), 145-162.

LeGendre, S. J., Liss, J. M., Lotto, A. J., & Utianski, R. (2009). Talker recognition using envelope modulation spectra. *Poster presentation at the Acoustical Society of America, San Antonio, Tx*.

Lehiste, I. (1977). Isochrony Reconsidered. *Journal of Phonetics*, 5, 253-263.

Lewandowski, N. (2011). *Talent in nonnative phonetic convergence*. Unpublished doctoral dissertation, University of Stuttgart.

Liss, J. M., Legendre, S., & Lotto, A. J. (2010). Discriminating dysarthria type from envelope modulation spectra. *Journal of Speech, Language, and Hearing Research*, 53, 1246-1255.

Low, E. L., Grabe, E. & Nolan, F. (2000). Quantitative characterizations of speech rhythm: syllable-timing in Singapore English. *Language & Speech*, 43, 377-401.

Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research, 50, 940-967*.

MATLAB (2012b). The MathWorks, Inc., Natick, MA.

Morrison, G. S. & Nearey, T. M. (2007). Testing theories of vowel inherent spectral change. *Journal of the Acoustical Society of America*, 122(1), EL15-EL22.

Namy, L. L., Nygaard, L. C. & Sauerteig, D. 2002. Gender differences in vocal accommodation: The role of perception. *Journal of Language and Social Psychology*, 21(4), 422–432.

Nearey, T.M. and Assmann, P.F. (1986). Modeling the role of inherent spectral change in vowel identification. *Journal of the Acoustical Society of America*, 80, 1297-1308.

Ní Chiosáin, M. (2007). Effects of synchronous speech task on length and prosody. In interdialectal nonprestige varieties. *Language Variation and Change*, 19, 51–62.

Nielsen, K. Y. (2008). Word-level and Feature-level Effects in Phonetic Imitation. Unpublished doctoral dissertation, University of California, Los Angeles.

Nenkova, A., Gravano, A. & Hirschberg, J. (2008). High frequency word entrainment in spoken dialogue. *Proceedings of ACL/HLT*, 169-172.

O'Rourke, E. (2008a). Correlating Speech Rhythm in Spanish: Evidence from Two Peruvian Dialects. In J. Bruhn de Garavito and E. Valenzuela (Eds.), Selected Proceedings of the 10th Hispanic Linguistics Symposium (pp. 276-287). Somerville, MA: Cascadilla Proceedings Project.

O'Rourke, E. (2008b). Variability Index and Variation Coefficients to Peruvian Spanish. *Proceedings of Speech Prosody 2008: Fourth Conference on Speech Prosody*, 431–434.

Peterson, G. & Barney, H.L. (1952). Control methods used in a study of the vowels. *Journal of the Acoustical Society of America*, 24,175–184.

Pardo, J. S. (2006). On phonetic convergence during conversational interaction. *The Journal of the Acoustical Society of America*, *119*(4), 2382-2393.

Pardo, J. S., Jay, I. C., & Krauss, R. M. (2010). Conversational role influences speech imitation. *Attention, perception and psychophysics*, *72*(8), 2254–2264.

Pardo, J. S., Jordan, K., Mallari, R., Scanlon, C., & Lewandowski, E. (2013). Phonetic Convergence in Shadowed Speech: The Relation Between Acoustic and Perceptual Measures. *Journal of Memory & Language*, *69*, 183–195.

Patel, A. D. (2008). *Music, Language and the Brain*. , New York, NY: Oxford University Press.

Patel, A. D., Iversen, J. R., Bregman, M. R. & Schulz, I. (2009). Studying synchronization to a musical beat in nonhuman animals. *Annals of the New York Academy of Sciences*, 1169, 459-469.

Peretz, I. (2006). The nature of music from a biological perspective. *Cognition*, 100, 1–32.

Peterson, G. &Lehiste, I. (1960). Duration of syllable nuclei in English. *The Journal of the Acoustical Society of America*. 32, 693-703.

Pike, K. L. (1946). *The intonation of American English*. Ann Arbor: University of Michigan Press.

Pingali, S. (2009). *Dialects of English: Indian English*. Scotland, UK: Edinburgh University Press.

R Core Team (2012). R: A language and environment for statistical computing. http://www.R-project.org/

Ramus, F., Dupoux, E., & Mehler, J. (2003). The psychological reality of rhythm classes: Perceptual studies. Paper presented at the *15th International Congress of Phonetic Sciences*, Barcelona, 337-342.

Ramus, F. & Mehler, J. (1999). Language identification with supra-segmental cues: A study based on speech resynthesis. *Journal of the acoustical Society of America*, *5(1), 512-521*.

Ramus, F., Hauser, M. D., Miller, C., Morris, D., & Mehler, J. (2000). Language Discrimination by Human Newborns and by Cotton-Top Tamarin Monkeys. *Science, 288, 26-35*.

Ramus, F., Nespor, M. & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, 73, 265-292.

Rao, G., & Smiljanic, R. (2011a). Effect of language, speaking style and talker on a spectral rhythm measure. *The Journal of the Acoustical Society of America*, *129*, 2680.

Rao, G. & Smiljanic, R. (2011b). Effects of Language, Speaking Style and Age on Prosodic Rhythm. *Proceedings of the XVIIth international Congress of Phonetic Sciences*, Hong Kong, China.

Rao, G., Smiljanic, R., & Diehl. R (2011c). Effect of linguistic background on convergence of prosodic rhythm. *The Journal of the Acoustical Society of America*, *133*, 3340.

Shepard, C. A., Giles, H. & LePoire, B. (2001). Communication Accommodation Theory. In The New Handbook of Language and Social Psychology (Robinson W. P. and Giles, H. (Ed.). Chichester: Wiley.

Sancier, M. L. & Fowler, C. A. (1997). Gestural drift in a bilingual speaker of Brazilian Portuguese and English. *Journal of Phonetics*, 25, 412-436.

Skowronski, M. D. & Harris, J. G. (2006). Applied principles of clear and Lombard speech for intelligibility enhancement in noisy environments. *Speech Communication*, 48 (5), 549-558.

Smiljanic, R. & Bradlow, A. R. (2008). Temporal organization of English clear and plain speech. *Journal of the Acoustical Society of America*, 124(5), 3171-3182.

Tilsen, S. & Johnson, K. (2008). Low-frequency Fourier analysis of speech rhythm. *Journal of the Acoustical Society of America*, 124(2), EL34-EL39.

Tomasello, M. (2009). Acceptance speech of the Hegel Prize. Retrieved from http://www.stuttgart.de/img/mdb/item/383875/51641.pdf

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: the origins of cultural cognition. *The Behavioral and brain sciences*, *28*(5), 675–91; discussion 691–735.

Warneken, F., & Tomasello, M. (2006). Altruistic helping in human infants and young chimpanzees. *Science*, *311*(5765), 1301–1303.

White, L. & Mattys S. L. (2007). Calibrating rhythm: First language and second language studies. *Journal of Phonetics*, 35, 501–522.

Wiget, L., White, L., Schuppler, B., Grenon, I., Rauch, O. & Mattys, S. L. (2010). How stable are acoustic metrics of contrastive speech rhythm? *The Journal of the Acoustical Society of America*, 27, 1559-1569.

White, L., Mattys, S. L., Grenon, I., & Wiget, L. (2007). That elusive rhythm : Pros and cons of rhythm metrics, *Laboratory phonology*, 11, 161–162.

Wiltshire, C. R., & Harnsberger, J. D. (2006). The influence of Gujarati and Tamil L1s on Indian English: a preliminary study. *World Englishes*, *25*(1), 91–104.

Vita

Gayatree Rao was born and raised in India. She has a Bachelor of Science in Engineering (Computer Engineering) degree from the University of Michigan in Ann Arbor and a Masters of Arts (Linguistics) from University of Texas – Austin.

Permanent email address: raog@utexas.edu

This dissertation was typed by the author.